INVESTIGATION OF READING AND WRITING GROWTH USING CURRICULUM BASED MEASUREMENT: SCHOOL DISTRICT 57 (PRINCE GEORGE)

by

Christopher S. Gilbert

B.A., Simon Fraser University, 1988
P.D.P., Simon Fraser University, 1991

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION

in

EDUCATIONAL COUNSELLING

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

December 2005

ABSTRACT

This study investigated two Curriculum Based Measurement data sets with the intent of comparing norm and gender differences over time. The original samples consisted of elementary students randomly selected for the Prince George School District norming project. A sample group of 2200 students was used in 1995/96 and a sample group of 2225 was used in 2002/03. The measurements were collected by teachers and other school district staff in each elementary school during October, January, and April. Only the April 1996 and April 2003 scores were used in this study. Descriptive and effect sizes for Grades 1 to 7 for the April testing period was generated for CBM measures entitled Words Read Correctly, Words Spelled Correctly, and Total Words Written. Statistical inferences were made based on plots of 95%confidence intervals placed on performance by grade line graphs of groups, years, and genders. The analyses were repeated at the tenth percentile mark. In comparing the results, this study documented that females continue to score significantly higher than males in reading, writing, and spelling, although males have narrowed the gap particularly in the intermediate grades. Males have significantly made the greatest gains in reading at almost every grade level and both genders have made some gains at the 10^{th} percentile, with males showing the most growth. The comparative data highlights the necessity of renorming studies taking place on a regular basis and the continued search for strategies to reduce the differences in gender scoring.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	viii
Acknowledgement	x
CHAPTER ONE: INTRODUCTION	1
CHAPTER TWO: METHOD	4
Participants	4
Instruments	7
Procedures	8
Ethics	9
CHAPTER THREE: RESULTS	10
SERIES ONE: Introduction of Combined Genders 2003 versus 1999	10
Words Read Correctly – Combined Genders– 2003 vs. 1996	11
Words Spelled Correctly – Combined Genders– 2003 vs. 1996	13
Total Words Written – Combined Genders– 2003 vs. 1996	16
Summary of Combined Genders 2003 versus 1996	19
SERIES TWO: Introduction to Females versus Males – 1996 and 2003	20
Words Read Correctly – Females vs. Males – 1996	20
Words Read Correctly – Females vs. Males – 2003	23
Words Spelled Correctly – Females vs. Males – 1996	26
Words Spelled Correctly – Females vs. Males – 2003	29
Total Words Written – Females vs. Males – 1996	32
Total Words Written – Females vs. Males – 2003	35
Summary of Females versus Males 1996 and 2003	38
SERIES THREE: Introduction of Female 2003 versus Female 1996 and Male 2003 versus Male 1996	39
Words Read Correctly – Females – 2003 vs. 1996	39
Words Read Correctly – Males – 2003 vs. 1996	42
Words Spelled Correctly – Females – 2003 vs. 1996	45
Words Spelled Correctly – Males – 2003 vs. 1996	48
Total Words Written – Females – 2003 vs. 1996	51
Total Words Written – Males – 2003 vs. 1996	54
Summary of Female 2003 versus Female 1996 and Male 2003 versus Male 1996	57

SERIES FOUR: Introduction of 10th Percentile – 2003 versus 1996 (Combined Genders/Female vs. Female/Male vs. Male) 58
Words Read Correctly – 10th Percentile – Combined Genders – 2003 vs. 1996 58
Words Spelled Correctly – 10th Percentile – Combined Genders – 2003 vs. 1996 61
Total Words Written – 10th Percentile – Combined Genders – 2003 vs. 1996 64
Words Read Correctly – 10th Percentile – Females – 2003 vs. 1996 67
Words Read Correctly – 10th Percentile – Males – 2003 vs. 1996 70
Words Spelled Correctly – 10th Percentile – Females – 2003 vs. 1996 73
Words Spelled Correctly – 10th Percentile – Males – 2003 vs. 1996 76
Total Words Written – 10th Percentile – Females – 2003 vs. 1996 79
Total Words Written – 10th Percentile – Males – 2003 vs. 1996 82
Summary of 10th Percentile – 2003 versus 1996 (Combined Genders/Female vs. Female/Male vs. Male) 85
CHAPTER FOUR: DISCUSSION AND CONCLUSION 87
Limitation of the study 88
REFERENCES 89

LIST OF TABLES

Table 1: Total Number of Students per Grade for each CBM Variable in April 1996: Words Read Correctly (WRC); Words Spelled Correctly (WSC); Total Words Written (TWW) 6

Table 2: Total Number of Students per Grade for each CBM Variable in April 2003: Words Read Correctly (WRC); Words Spelled Correctly (WSC); Total Words Written (TWW) 7

Table 3: Descriptive Statistics for WRC – Combined Genders – 2003 vs.1996 11

Table 4: Effect Sizes for WRC – Combined Genders – 2003 vs. 1996 13

Table 5: Descriptive Statistics for WSC – Combined Genders – 2003 vs.1996 14

Table 6: Effect Sizes for WSC – Combined Genders – 2003 vs. 1996 16

Table 7: Descriptive Statistics for TWW – Combined Genders – 2003 vs. 1996 17

Table 8: Effect Sizes for TWW – Combined Genders – 2003 vs. 1996 19

Table 9: Descriptive Statistics for WRC – Females vs. Males – 1996 21

Table 10: Effect Sizes for WRC – Females vs. Males – 1996 23

Table 11: Descriptive Statistics for WRC – Females vs. Males – 2003 24

Table 12: Effect Sizes for WRC – Females vs. Males – 2003 26

Table 13: Descriptive Statistics for WSC – Females vs. Males – 1996 27

Table 14: Effect Sizes for WSC – Females vs. Males – 1996 29

Table 15: Descriptive Statistics for WSC – Females vs. Males – 2003 30

Table 16: Effect Sizes for WSC – Females vs. Males – 2003 32

Table 17: Descriptive Statistics for TWW – Females vs. Males – 1996 33

Table 18: Effect Sizes for TWW – Females vs. Males – 1996 35

Table 19: Descriptive Statistics for TWW – Females vs. Males – 2003 36

Table 20: Effect Sizes for TWW – Females vs. Males – 2003 38

Table 21: Descriptive Statistics for WRC – Females – 2003 vs. 1996 40

Table 22: Effect Sizes for WRC – Females – 2003 vs. 1996 42

Table 23: Descriptive Statistics for WRC – Males – 2003 vs. 1996 43

Table 24: Effect Sizes for WRC – Males – 2003 vs.1996 45

Table 25: Descriptive Statistics for WSC – Females – 2003 vs.1996 46

Table 26: Effect Sizes for WSC – Females – 2003 vs.1996 48

Table 27: Descriptive Statistics for WSC – Males – 2003 vs.1996 49

Table 28: Effect Sizes for WSC – Males – 2003 vs.1996 53

Table 29: Descriptive Statistics for TWW – Females – 2003 vs.1996 52

Table 30: Effect Sizes for TWW – Females – 2003 vs.1996 54

Table 31: Descriptive Statistics for TWW – Males – 2003 vs.1996 55

Table 32: Effect Sizes for TWW – Males – 2003 vs.1996 57

Table 33: Descriptive Statistics for WRC – 10^{th} Percentile – Combined Genders – 2003 vs.1996 59

Table 34: Effect sizes for WRC – 10thPercentile – Combined Genders – 2003 vs.1996 61

Table 35: Descriptive Statistics for WSC – 10^{th} Percentile – Combined Genders – 2003 vs.1996 62

Table 36: Effect Sizes for WSC – 10^{th} Percentile – Combined Genders – 2003 vs.1996 64

Table 37: Descriptive Statistics for TWW – 10^{th} Percentile – Combined Genders – 2003 vs.1996 65

Table 38: Effect sizes for TWW – 10^{th} Percentile – Combined Genders – 2003 vs.1996 67

Table 39: Descriptive Statistics for WRC – 10th Percentile – Females – 2003 vs.1996 68

Table 40: Effect sizes for WRC – 10th Percentile – Females – 2003 vs.1996 70

Table 41: Descriptive Statistics for WRC – 10th Percentile – Males – 2003 vs.1996 71

Table 42: Effect sizes for WRC – 10th Percentile – Males – 2003 vs.1996 73

Table 43: Descriptive Statistics for WSC – 10th Percentile – Females –2003 vs.1996 74

Table 44: Effect sizes for WSC – 10th Percentile – Females – 2003 vs.1996 76

Table 45: Descriptive Statistics for WSC – 10th Percentile – Males – 2003 vs.1996 77

Table 46: Effect sizes for WSC – 10th Percentile – Males – 2003 vs.1996 79

Table 47: Descriptive Statistics for TWW – 10th Percentile – Females – 2003 vs.1996 80

Table 48: Effect sizes for TWW – 10th Percentile – Females – 2003 vs.1996 82

Table 49: Descriptive Statistics for TWW – 10th Percentile – Males – 2003 vs.1996 83

Table 50: Effect sizes for TWW – 10th Percentile – Males – 2003 vs.1996 85

LIST OF FIGURES

Figure 1: Grade Score Means in WRC for Combined Genders – 2003 vs. 1996 12

Figure 2: Grade Score Means in WSC for Combined Genders – 2003 vs. 1996 15

Figure 3: Grade Score Means in TWW for Combined Genders – 2003 vs. 1996 18

Figure 4: Grade Score Means for WRC – Females vs. Males – 1996 22

Figure 5: Grade Score Means for WRC – Females vs. Males – 2003 25

Figure 6: Grade Score Means for WSC – Females vs. Males – 1996 28

Figure 7: Grade Score Means for WSC – Females vs. Males – 2003 31

Figure 8: Grade Score Means for TWW – Females vs. Males – 1996 34

Figure 9: Grade Score Means for TWW – Females vs. Males – 2003 37

Figure 10: Grade Score Means for WRC – Females – 2003 vs. 1996 41

Figure 11: Grade Score Means for WRC – Males – 2003 vs. 1996 44

Figure 12: Grade Score Means for WSC – Females – 2003 vs. 1996 47

Figure 13: Grade Score Means for WSC – Males – 2003 vs. 1996 50

Figure 14: Grade Score Means for TWW – Females – 2003 vs. 1996 53

Figure 15: Grade Score Means for TWW – Males – 2003 vs. 1996 56

Figure 16: Grade Score Means in WRC – 10^{th} Percentile – Combined Genders – 2003 vs. 1996 60

Figure 17: Grade Score Means in WSC – 10^{th} Percentile – Combined Genders – 2003 vs. 1996 63

Figure 18: Grade Score Means in TWW – 10^{th} Percentile – Combined Genders – 2003 vs. 1996 66

Figure 19: Grade Score Means for WRC – 10^{th} Percentile – Females – 2003 vs. 1996 69

Figure 20: Grade Score Means for WRC – 10th Percentile – Males – 2003 vs. 1996 72

Figure 21: Grade Score Means for WSC – 10th Percentile – Females – 2003 vs. 1996 75

Figure 22: Grade Score Means for WSC – 10th Percentile – Males – 2003 vs. 1996 78

Figure 23: Grade Score Means for TWW – 10th Percentile – Females – 2003 vs. 1996 81

Figure 24: Grade Score Means for TWW – 10th Percentile – Males – 2003 vs. 1996 84

ACKNOWLEDGEMENT

The journey towards this masters degree started with recommendations from Al Curtis, Bruce Ballantyne, Brian Toll, and Robin Temoin. I thank these gentlemen for their support before the program began.

I'm grateful to past and present UNBC teaching staff, in particular: Dr. Colleen Haney, who was the first to welcome me to the program and exposed me to new counseling theories; Dr. Ron Lehr, who observed my early counseling sessions behind a one-way mirror and then provided the challenge of a counseling practicum at UNBC; Dr. Bryan Hartman, who introduced career counseling and was key in getting the degree completion process back on track; Dr. Glen Schmidt, who graciously agreed to be one of my committee members; and last but certainly not least, Dr. Peter MacMillan, who I began the program with in my first course and who has been instrumental in guiding me home through the last stage of degree completion.

I'm also appreciative for the work of Lynn Hedekar, Shelley Wiltshire, John Cook, and Saima Fewster. Their research provided an integral base for this study.

Family is everything and I'm blessed to have a strong supporting cast, which includes: my mother Katherine, who originally encouraged me to pursue post secondary education; my father Henry, who has taught me to think outside the box; and my eleven siblings (Elizabeth, Timothy, William, Paul, Rebecca, Robert, Sarah, Mark, Matthew, Tristan, and Megan), who are a joy to be around and with them there is never a dull moment.

I would like to dedicate this project to my late-grandmother (Nana) and my wife Rhonda. Nana, with her innate curiosity, gave me the impetus to venture beyond the routine and seek adventure. Rhonda, my partner for over two decades, has seen me at my best and worst, yet has never wavered in her encouragement. Her patience, love, and support greatly contributed to me seeing this degree to its eventual conclusion.

CHAPTER ONE: INTRODUCTION

The purpose of this study is to compare the results of two established and validated Curriculum Based Measurement (CBM) data sets, one of which was collected in 1995/96 and the other collected in 2002/03. This study will document changes in the norms of elementary school students from each data set and also compare gender differences in reading fluency and written expression fluency as measured by CBM variables. Discussion will reflect upon whether differences are consistent throughout the elementary grades and whether the gap between the genders has increased, decreased, or remained the same over the two time periods.

Information from Hedekar's 1997 thesis, Wiltshire's 2003 thesis, and Cook's 2005 project will be cited in the method chapter due to their analysis of the CBM data sets. Sources that these authors use that are in this project will be included in the reference list. Given the nature of this project, no new references are cited.

Hedekar's 1997 thesis used the 1995/96 CBM data set to investigate the relationship between scores on achievement measures, the gender of the students, and the month in which they were born. Hedekar's literature review found a consistent gender difference, with males scoring lower than females in 12 of 16 studies. No gender difference was discovered in 2 studies while 2 other studies did not measure gender differences (p.17). Hedekar (1997) found a gender difference favoring females in all the analyses for words spelled correctly (WSC) and total words written (TWW) for Grades 1 through 7. A gender difference favoring females was also found in 14 of 19 analyses for words read correctly (WRC) for Grades 1 through 7 in the same report.

Wiltshire's 2003 thesis also investigated gender differences in the analysis of the 2002/03 data set. Wiltshire's literature review included Gambell and Hunter's (2000) survey on gender differences in Canada. In the survey females' outperformed males in provincial exam results in all literacy based courses such as English, French, Communications, and Literature (Wiltshire, 2003, p.19). In analyzing Allred's (1990) study on spelling ability, where females in Grade 1 through 6 significantly outscored males, Wiltshire asserts that gender differences in spelling relate to gender differences in reading achievement (Wiltshire, 2003, p. 28).

Wiltshire (2003) found that female students scored higher in literacy skills in reading and writing fluency at every grade level and every testing period, with the consistent statistically significant gender difference in writing fluency as measured by WSC and TWW from Grade 1 through 7 (p. 86).

Wiltshire did not investigate relative-age within grade differences even though the data would have permitted such analyses due to the lack of evidence of any effect found in Hedekar's (1997) study. Wiltshire did report on differences and trends between aboriginal students and others, but Hedekar did not, thus a comparison was not possible between the two data sets.

Cook's 2005 project analysis, based on the 2002/2003 DIBELS/CBM data set, indicated that the CBM measures were reliable and possessed the technical qualities necessary for use as intended by School District 57. Cook (2005) found a high correlation between the three testing times of October, January, and April in the 1995/96 and 2002/03 data sets. The reliability in the measures asserted by Cook's 2005 project justifies the decision to use and compare only April scores for this project. Cook (2005)

concludes that "the CBM data set can be considered to be of good quality and of high reliability because of the equivalency of the probes used in both reading and writing, the normal distributions, high correlations between testing periods, stable validity correlations and good to excellent Cronbach alpha coefficients" (p.32).

Fewster & MacMillan (2002) explored the relationship between Grades 6 and 7 CBM oral reading fluency and written expression scores with their year-end English and social studies marks received in Grades 8, 9, and 10. Their study found correlations between WRC scores and WSC scores in Grade 6 and 7 with school-awarded grades in English and social studies in Grades 8, 9, and 10 were significant and generally of medium effect size. In most cases WRC correlated more highly with course grades than WSC. At every grade level, both the WRC and WSC measures correlated more highly with English than with social studies scores. Correlations were most often highest in grade 8 and gradually decreased over time.

Fewster & MacMillan (2002) also explored the predictability of junior secondary students (Grades 8-10) membership into special education classes, learning assistance classes, general education classes, and honors classes from their elementary school CBM scores. The results found all comparisons between the four program placement groups for both WRC and WSC being significant at the $p < .0005$ level.

Despite the limitation of only using Grade 6 and 7 CBM data and in some cases, linking only to junior secondary school work, CBM ability to predict academic success and program placement makes it a highly useful indicator for use in School District 57.

CHAPTER TWO: METHOD

This chapter contains four sections. The first section describes the participants who were tested and how they were selected for CBM norming for each project. The second section describes the test instruments used for the CBM norming project of each study and the advantages for using them. The third section is a description of the procedures used for the studies. The fourth section discusses ethics.

Participants

The CBM scores of students in the 1995/96 CBM data set were collected by teachers within School District 57 (Prince George). The students were randomly chosen to participate in the CBM testing at each school grade based on the random numbers generated by the project coordinators employing a sample stratified by school (Hedekar, 1997). Over twenty percent of the students in the school district's elementary schools were tested in October, January, and April during the 1995/96 year (as only the April administrative dates included Grade One students for both data sets, only April scores will be reported in this project). The 1995/96 data set had 2367 students (including kindergarten) tested for the norming project. Randomly selected students who later transferred out of a school attendance area were replaced by other students who had recently moved into the school. This enabled the norming project to retain the full 20% at each testing period without biasing the sample with smaller numbers of transient students (Hedekar, 1997).

Due to the nature of the 1995/96 study, where not only gender differences were to be analyzed but also birth month, a number of students were excluded in the data analysis. Students who were not at the appropriate age for their grade level were

excluded. These included students who had enrolled in school a year later than their age mates, students who had been retained for a second year at any of the grade levels, and students who had been accelerated to a higher grade level. Students who transferred in or out of the study were also excluded from the study. The 1995/96 data set did not include level one and level two ESL students, students with mental disabilities, other hard-labeled students such as hearing-impaired, visually impaired, autistic students, and students enrolled in French Immersion. The 2002/03 data collection also used these exclusion rules.

Cook (2005) states that sample size is very important when developing district norms. His source (Shinn, 1989) suggests that 100 to 150 students per grade are needed in the sample group to develop district norms. The results section of this report will show that samples range from 104 to 335 students. The number of students in each grade level analyzed in April 1996 for each measure is provided in Table 1. The number of students in Grade One, Two, and Three in 1996 are lower for the Words Spelled Correctly (WSC) data, as reporting WSC was optional for primary students. The reason for this was the difference among teachers in strategies for teaching spelling and phonics at the lower grade levels (Hedekar, 1997). This difference appears to have been negligible in the second data set, as collecting data for Words Spelled Correctly was expected in 2002/03.

Table 1

Total Number of Students per Grade for each CBM Variable in April 1996: Words Read Correctly (WRC); Words Spelled Correctly (WSC); Total Words Written (TWW)

GRADE	WRC	WSC	TWW
1	318	224	318
2	314	215	312
3	320	234	319
4	318	318	318
5	311	309	314
6	320	317	317
7	319	320	320
TOTAL	2220	1937	2218

The data set for 2002/03 students' reading and writing fluency followed the same method of selecting students as outlined in the 1995/96 collection. The 2002/02 data set consisted of 2272 elementary students (including kindergarten) randomly selected within School District 57 (Prince George). The number of students in each grade level analyzed for each measure is provided in Table 2.

Table 2

Total Number of Students per Grade for each CBM Variable in April 2003: Words Read Correctly (WRC); Words Spelled Correctly (WSC); Total Words Written (TWW)

GRADE	WRC	WSC	TWW
1	247	247	247
2	265	265	265
3	281	279	279
4	309	309	309
5	276	277	277
6	312	311	311
7	335	334	334
TOTAL	2025	2022	2022

Instruments

Curriculum Based Measurement (CBM) is a specific approach for measuring students' growth in basic skills (Hedekar, 1997). This measurement approach was developed by the Minnesota Institute for Research on Learning Disabilities during a six year period starting in 1977 (Deno, 1992 in Hedekar, 1997). The original research focused on characteristics of students referred and pronounced eligible for special education as well as instructional interventions and evaluations (Ysseldyke, et al., 1983 in Hedekar, 1997). CBM was developed to be a measurement approach that would be efficient, valid, reliable, easily understood by teachers, and inexpensive to administer

(Deno, 1985 in Hedekar, 1997). Nowadays CBM is one tool used in determining eligibility for additional services within the school. Studies show that CBM is as effective as standardized testing for determining students' eligibility for support (Dunn, 1991; Shinn, Nolet, & Knutson, 1990 in Hedekar, 1997).

The key components of CBM are probes, short concise measurement tests designed to assess fluency. In each study, a reading probe consisted of a reading passage from a prescribed grade level reading textbook used in the school district. The reading materials selected for the CBM reading probes represented a sample of reading curriculum materials used in the daily teaching of reading to students in the school district (Hedekar, 1997, p. 23). The measure of oral reading was the number of words that a student was able to read aloud correctly (WRC) in one minute from a grade level reading probe. A writing probe consists of a story starter sentence from which the student was asked to write a story. The writing measure is taken from a three minute timed sample. Two scores were calculated. One score was the number of words written (TWW); the second was words spelled correctly (WSC) (Cook, 2005).

Procedures

Students at each grade were assessed on the grade level materials for the grade in which they were enrolled. The first grade students were given reading probes designed from first grade reading materials, all second grade students were given second grade reading probes and so on through the grades. This matching of grade level probes to students' enrolled grade continued for all grades regardless of what the learning assistance teacher knew about the students' actual reading level or ability (Hedekar, p. 20). Students were tested individually. At the appropriate testing time, the student left the

classroom and traveled to the testing area. The timed oral reading test was administered first. The student then received a designated written expression probe. The written expression tests were group administered in class settings. The six expression probes were used across all grades. The students were given a "story starter" then were instructed to think about their stories while sitting quietly for one minute. Then the students wrote for three minutes. The students selected for the CBM study had their written expression work scored by the special education teacher. Thus, the written expression performance of the students participating in the study was a sample of typical classroom work. The teachers administering and marking the tests had been trained during a series of district in-service workshops (Hedekar & MacMillan, p. 12, in progress).

Ethics

The Research Ethics Board chair commented that the application itself was unnecessary given the fact that this study was using intact data sets previously approved by the UNBC Ethics Review Committee.

CHAPTER THREE: RESULTS

The results for this project will be divided into four series of descriptive statistics, graph analysis, and effect size analysis for Words Read Correctly (WRC), Words Spelled Correctly (WSC), and Total Words Written (TWW). **Series one** will compare CBM results in 1996 and 2003 with full samples of all students (combined genders). This series will provide a full look at combined genders and will identify which grade means scores have increased for each CBM variable. **Series two** will compare females versus males on CBM scores at every grade level in 1996 and 2003. The task will be to determine if the gap between mean scores of females and males in 1996 has increased, remained the same, or decreased in 2003. **Series three** will compare female mean scores in 1996 with female mean scores in 2003 and male mean scores in 1996 with male mean scores in 2003. Changes in mean scores with each gender in each grade will be explored. **Series four** will explore mean score differences at the 10^{th} percentile for 1996 and 2003. Full samples of all students (combined genders) as well as each gender against itself in 1996 versus 2003 will be analyzed. A cut off percentile is often used as a gauge as to which students should receive special support within schools. A cut off around the 10^{th} percentile is sometimes used. Therefore, this series of samples will explore changes at the 10^{th} percentile between the two data sets.

SERIES ONE: Introduction of Combined Genders – 2003 versus 1996

The first series is a comparison of 1996 and 2003 full samples of all students' (combined genders) mean scores in WRC, WSC, and TWW in Grades 1 to 7

Words Read Correctly – Combined Genders– 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 3 for Words Read Correctly (WRC) for Combined Genders in 2003 and 1996. The largest sample size in 2003 was 335 in Grade 7 (see column 4, row 9) and the smallest sample size was 247 in Grade 1 (see column 4, row 3). In 1996 the largest sample size was 320 in Grade 3 and Grade 6 (see column 8, row 5 & column 8, row 8) and the smallest sample size was 311 in Grade 5 (see column 8, row 7). The mean score increased for each successive grade with both data sets. The 1996 median S.E. 2.30 multiplied by 1.96, provided the 95% confidence interval of ± 4.50 used in Figure 1.

Table 3

Descriptive Statistics for WRC – Combined Genders – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	36.02	29.60	247	1.88	36.66	31.94	318	1.79
Two	81.03	42.32	265	2.60	73.53	35.45	314	2.00
Three	110.31	39.47	281	2.35	103.40	41.13	320	2.30
Four	120.29	38.30	309	2.18	111.20	41.80	318	2.34
Five	130.57	38.55	276	2.32	123.80	41.86	311	2.37
Six	137.78	38.17	312	2.16	124.40	41.29	320	2.31
Seven	143.93	40.18	335	2.20	137.50	41.07	319	2.30

As illustrated in Figure 1, there is significant change ($p < .05$) in the mean scores of the two data sets for WRC Combined Genders. With the exception of Grade 1, all grades in 2003 significantly increased mean scores in comparison to the 1996 scores. As can be seen in Figure 1, the confidence intervals do not overlap for Grades 2 to 7, signifying growth between the two years. Grade 1 scores in 2003 were not significantly different than scores in 1996.

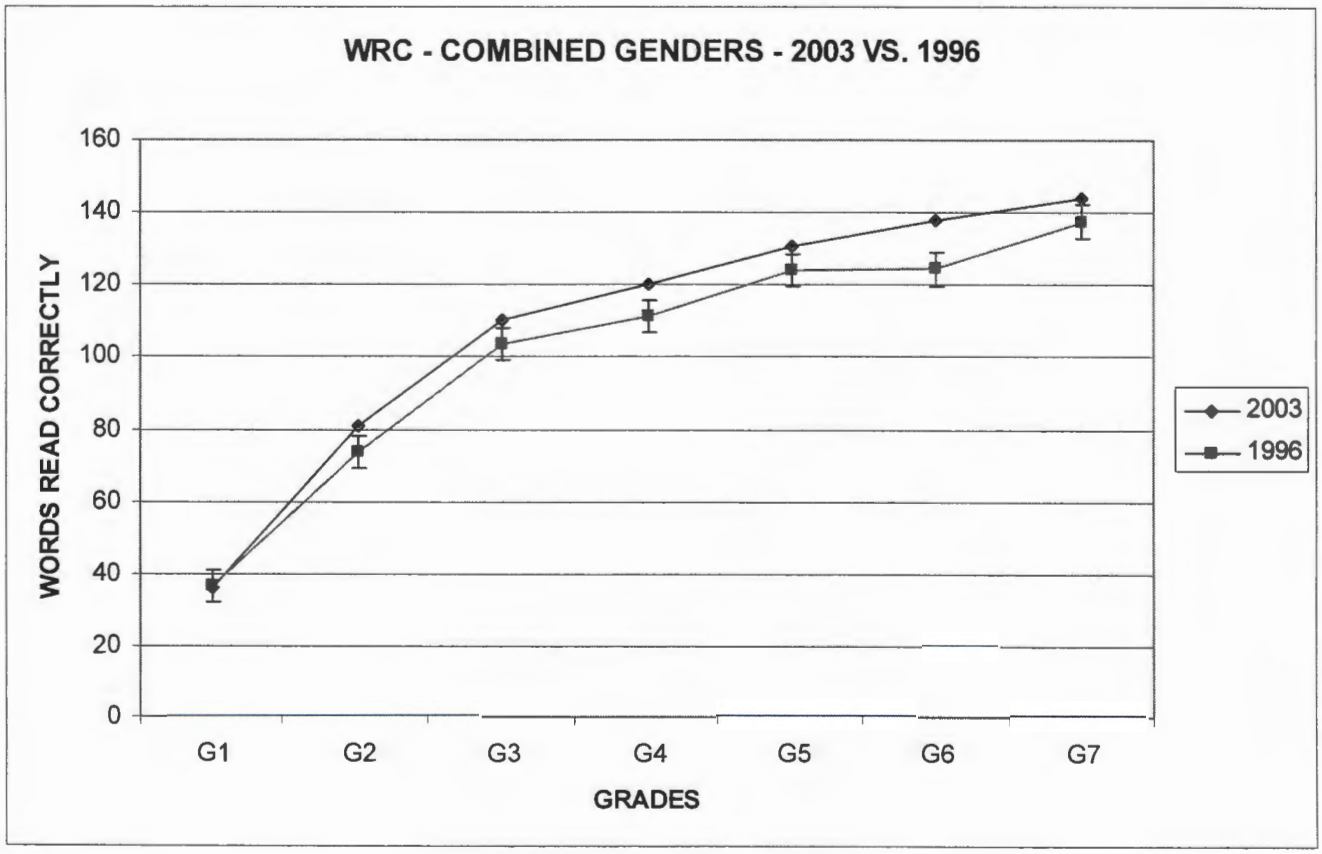

Figure 1. Grade Score Means in WRC for Combined Genders – 2003 vs. 1996

Cohen's *d* is calculated for WRC analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes where there was a significant difference of $p < .05$ are marked with an

asterisk. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 4, the effect size for each grade between the two data sets for WRC ranged from trivial to small.

Table 4

Effect Sizes for WRC – Combined Genders – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	247	36.02	318	36.66	31.94	0.00	trivial
Two	265	81.03	314	73.53	35.45	0.21	small *
Three	281	110.31	320	103.40	41.13	0.17	trivial *
Four	309	120.29	318	111.22	41.80	0.22	trivial *
Five	276	130.57	311	123.84	41.86	0.16	trivial *
Six	312	137.78	320	124.37	41.29	0.32	small *
Seven	335	143.93	319	137.50	41.07	0.16	trivial *

Words Spelled Correctly – Combined Genders– 2003 vs. 1996

In Table 5 for Words Spelled Correctly (WSC) for Combined Genders in 2003 and 1996 the largest sample size in 2003 was 334 in Grade 7 (see column 4, row 9) and the smallest sample size was 247 in Grade 1 (see column 4, row 3). In 1996 the largest sample size was 320 in Grade 7 (see column 8, row 9) and the smallest sample size was 215 in Grade 2 (see column 8, row 4). The mean score increased for each successive

grade with both data sets. The 1996 median S.E. 0.85 multiplied by 1.96 provided the 95% confidence interval of ± 1.67 used in Figure 2.

Table 5

Descriptive Statistics for WSC – Combined Genders – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	9.77	7.03	247	0.45	8.97	6.33	224	0.42
Two	22.63	10.83	265	0.67	22.93	10.35	215	0.71
Three	31.72	12.25	279	0.73	33.01	12.32	234	0.81
Four	43.12	14.84	309	0.84	40.33	15.22	318	0.85
Five	49.17	15.74	277	0.95	48.91	16.70	309	0.95
Six	56.96	17.33	311	0.98	54.85	15.45	317	0.87
Seven	63.29	16.90	334	0.93	60.73	17.55	320	0.98

As illustrated in Figure 2, there is significant change ($p < .05$) in the mean scores of the two data sets for WSC Combined Genders in Grades 4, 6, and 7. As can be seen in Figure 2, the confidence intervals do not overlap for Grades 4, 6, and 7, signifying growth from 1996 to 2003. Of note is that the Grade 2 and 3 mean scores in 2003 regressed from the scores in 1996, with Grade 3 scores close to being significant.

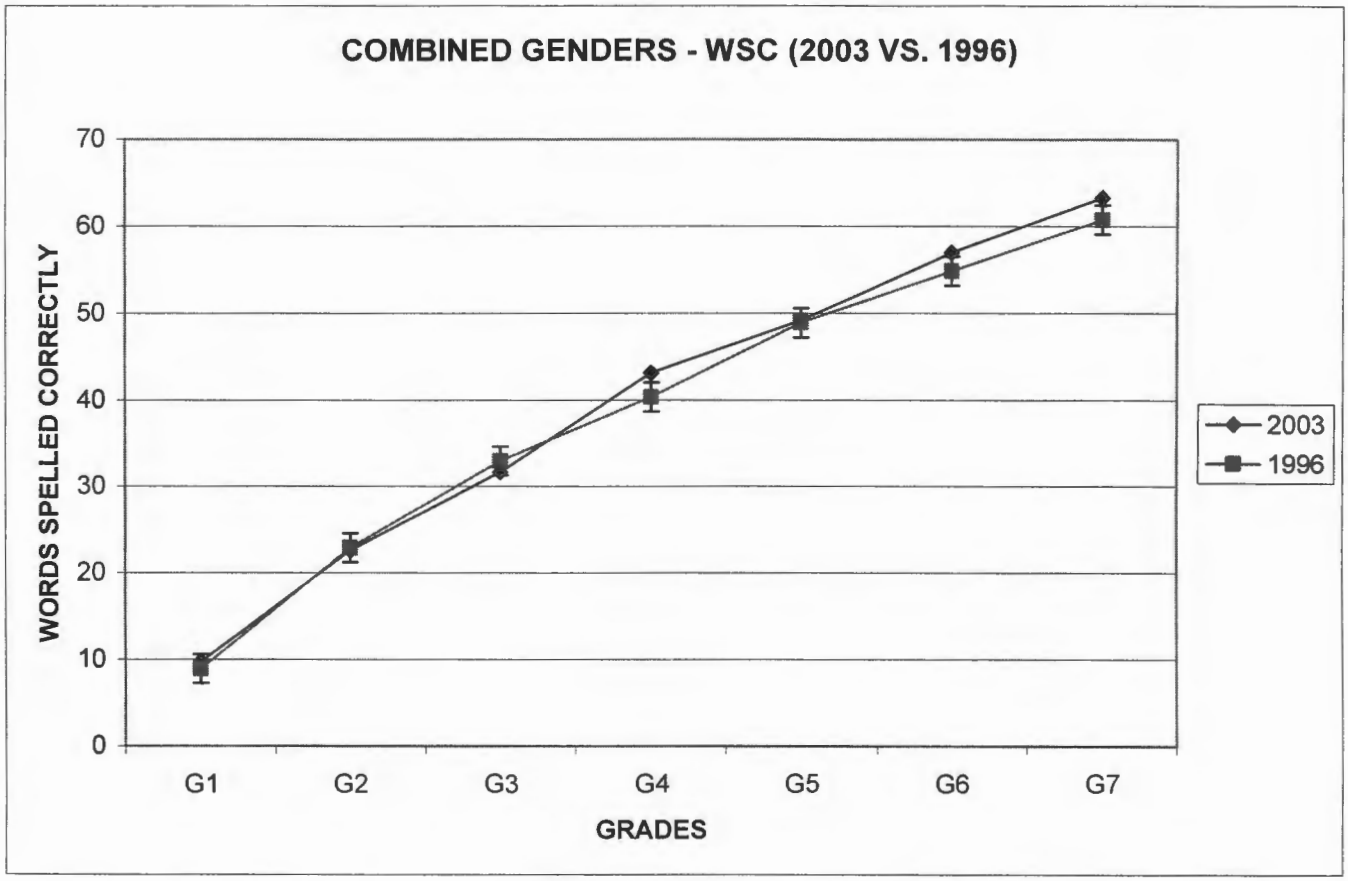

Figure 2. Grade Score Means in WSC for Combined Genders – 2003 vs. 1996

Cohen's *d* is calculated for WSC analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes where there was a significant difference of $p < .05$ are marked with an asterisk. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 6, the effect size for each grade between the two data sets for WSC was trivial.

Table 6

Effect Sizes for WSC – Combined Genders – 2003 vs. 1996

	(2003)		(1996)		(1996)	
Grade	N	M	N	M	SD	Effect Size
One	247	9.77	224	8.97	6.33	0.13 trivial
Two	265	22.63	215	22.93	10.35	0.00 trivial
Three	279	31.72	234	33.01	12.32	- 0.10 trivial
Four	309	43.12	318	40.33	15.22	0.18 trivial *
Five	277	49.17	309	48.91	16.70	0.02 trivial
Six	311	56.96	317	54.85	15.45	0.14 trivial *
Seven	334	63.29	320	60.73	17.55	0.15 trivial *

Total Words Written – Combined Genders– 2003 vs. 1996

In Table 7 for Total Words Written (TWW) for Combined Genders in 2003 and 1996 the largest sample size in 2003 was 334 in Grade 7 (see column 4, row 9) and the smallest sample size was 247 in Grade 1 (see column 4, row 3). In 1996 the largest sample size was 320 in Grade 7 (see column 8, row 9) and the smallest sample size was 312 in Grade 2 (see column 8, row 4). The mean score increased for each successive grade with both data sets. The 1996 median S.E. 0.85 multiplied by 1.96 provided the 95% confidence interval of ± 1.67 used in Figure 3.

Table 7

Descriptive Statistics for TWW – Combined Genders – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	13.50	8.28	247	0.53	13.15	7.48	318	0.42
Two	26.80	10.98	265	0.67	26.19	10.42	312	0.59
Three	35.00	12.39	279	0.74	36.40	11.97	319	0.67
Four	46.00	15.00	309	0.85	43.56	15.21	318	0.85
Five	51.60	15.77	277	0.95	52.57	16.16	314	0.91
Six	59.10	17.19	311	0.97	57.58	15.48	317	0.87
Seven	65.40	16.77	334	0.92	63.13	17.62	320	0.99

As illustrated in Figure 3, there is significant change ($p < .05$) in the mean scores of the two data sets for TWW Combined Genders in Grades 4 and 7. Grades 4 and 7 in 2003 significantly increased their mean scores in TWW in comparison to the 1996 scores. As can be seen in Figure 3, the confidence intervals do not overlap for Grades 4 and 7, signifying growth between the two data sets. The increase in Grade 6 scores was just inside the confidence intervals and the drop in Grade 3 scores was also just short of being significant.

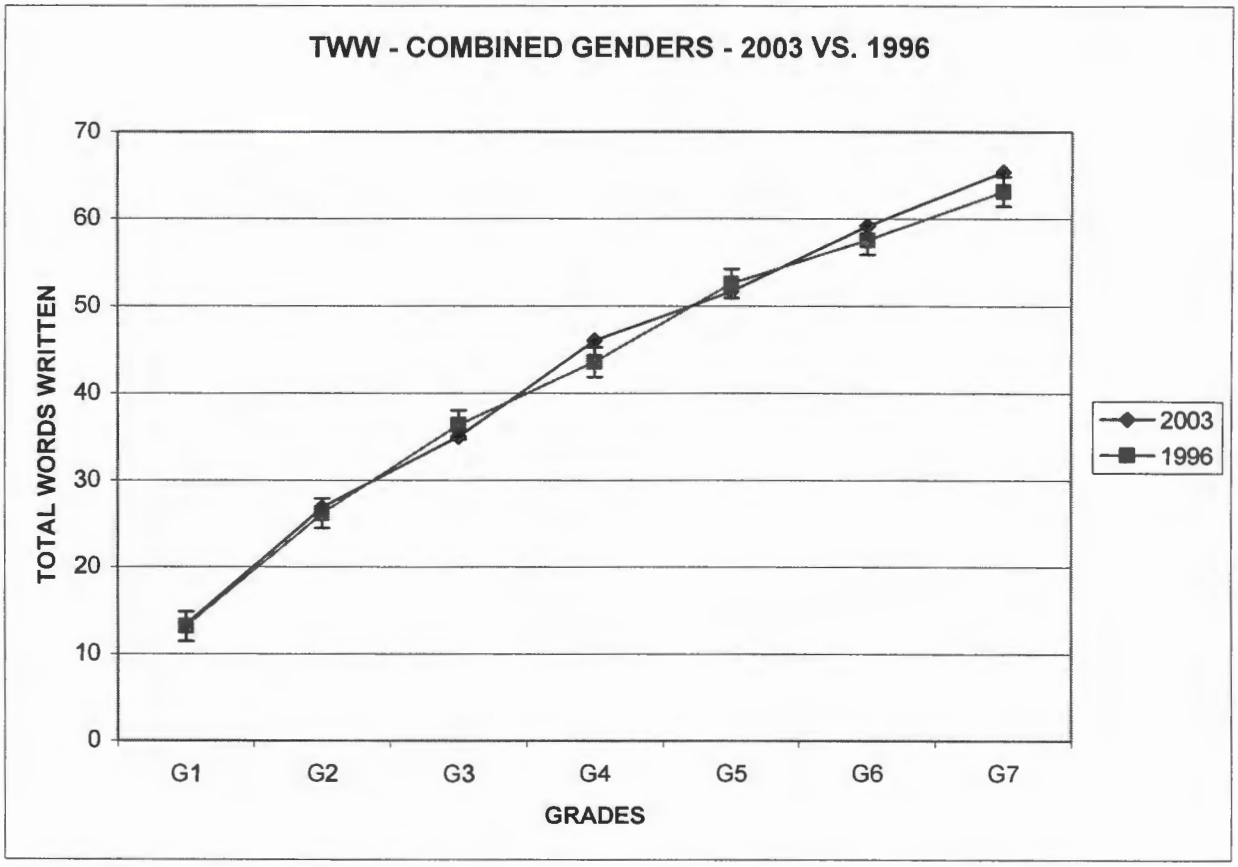

Figure 3. Grade Score Means in TWW for Combined Genders – 2003 vs. 1996

Cohen's *d* is calculated for TWW analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes where there was a significant difference of $p < .05$ are marked with an asterisk. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 8, the effect size for each grade between the two years for TWW was trivial.

Table 8

Effect Sizes for TWW – Combined Genders – 2003 vs.1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	247	13.45	318	13.15	7.48	0.04	trivial
Two	265	26.84	312	26.19	10.42	0.06	trivial
Three	279	35.01	319	36.40	11.97	- 0.10	trivial
Four	309	46.03	318	43.56	15.21	0.16	trivial *
Five	277	51.64	314	52.57	16.16	- 0.10	trivial
Six	311	59.14	317	57.58	15.48	0.10	trivial
Seven	334	65.40	320	63.13	17.62	0.13	trivial *

Summary of Combined Genders 2003 versus 1996

Based on the data it is apparent that the mean scores have generally increased for the intermediate grades (4 to 7) but not increased for the primary grades (1 to 3) in the three variables of WRC, WSC, and TWW. WRC had significant growth in 2003 for all grades except Grade 1 and the effect sizes ranged from trivial to small. WSC and TWW had significant growth in the intermediate grades but not at the primary levels and the effect sizes were all trivial. Both WSC and TWW had some grades that regressed in 2003.

SERIES TWO: Introduction to Females versus Males – 1996 and 2003

The second series will examine gaps between females and males in 2003 and 1996 and will determine whether the gaps between genders has increased, decreased, or remained the same in the two data sets.

Words Read Correctly – Females vs. Males - 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 9 for females versus males in Words Read Correctly (WRC) in 1996. The largest sample size for females was 174 in Grade 6 (see column 4, row 8) and the smallest sample size was 142 in Grade 3 (see column 4, row 5). The largest sample size for males was 178 in Grade 3 (see column 8, row 5) and the smallest sample size was 146 in Grade 6 (see column 8, row 8). The mean score increased for each successive grade with both data sets, with the exception of where females in Grade 6 scored less than females in Grade 5. The male median S.E. 3.04 multiplied by 1.96, provided the 95% confidence interval of ± 5.96 used in Figure 4.

Table 9

Descriptive Statistics for WRC – Females vs. Males – 1996

	(Females)				(Males)			
Grade	M	SD	N	SE	M	SD	N	SE
One	39.93	35.42	165	2.76	33.14	27.38	153	2.21
Two	82.60	34.45	159	2.73	64.21	34.12	155	2.74
Three	109.10	41.75	142	3.50	98.84	40.17	178	3.01
Four	116.20	43.03	163	3.37	105.96	39.94	155	3.21
Five	132.80	40.88	149	3.35	115.57	41.16	162	3.23
Six	130.20	39.90	174	3.03	117.41	41.97	146	3.47
Seven	145.90	42.31	164	3.30	128.60	37.86	155	3.04

As illustrated in Figure 4, there is significant difference ($p < .05$) in the mean scores of WRC Females versus Males in all Grades in 1996. As can be seen in Figure 4, the confidence intervals do not overlap for any of the grades, signifying a definite difference in scores between the two genders, with females being superior.

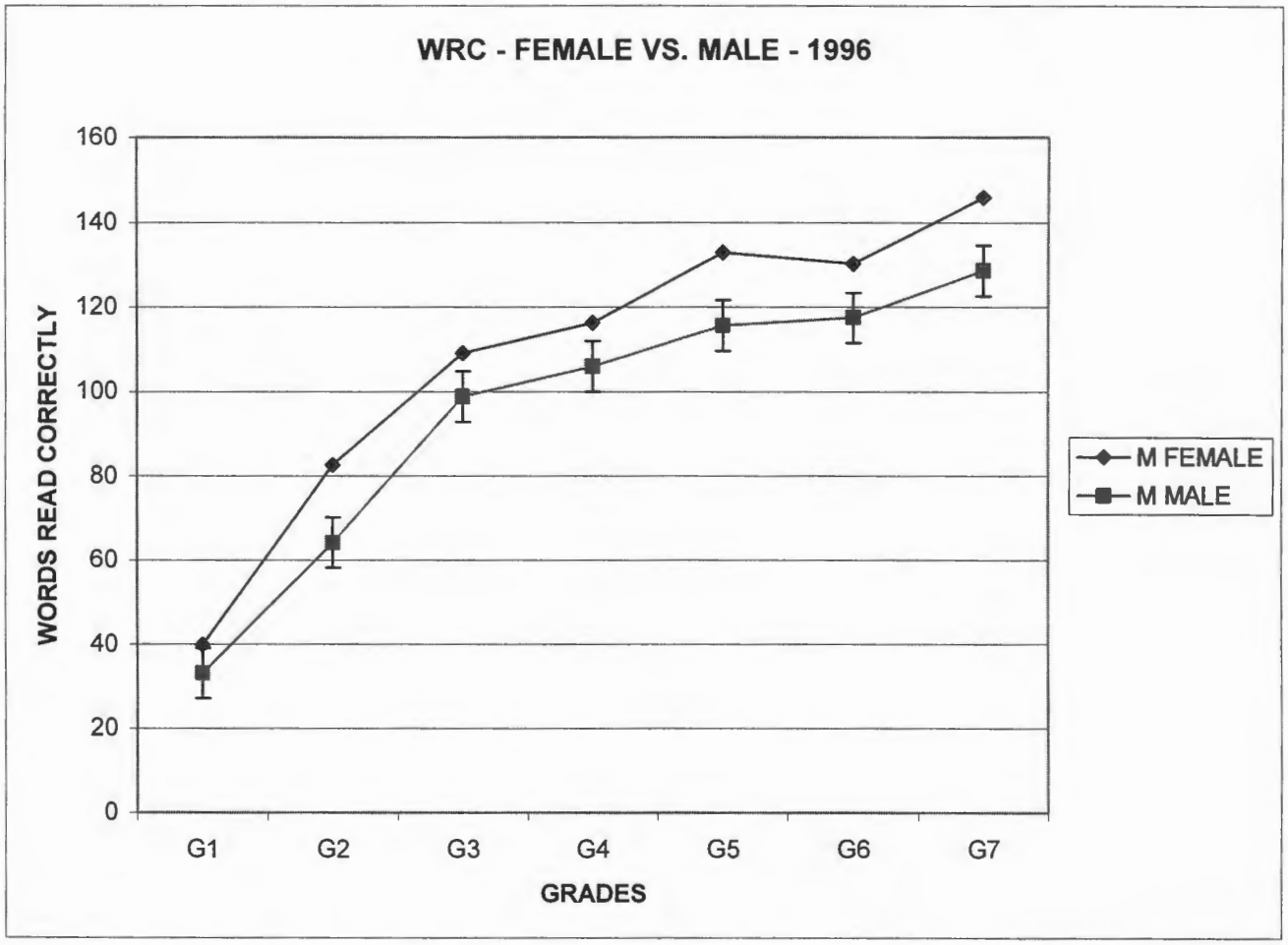

Figure 4. Grade Score Means for WRC – Females vs. Males – 1996

Cohen's *d* is calculated for WRC for each grade between each gender. The male standard deviation was used in the Cohen's *d* calculations. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 10, the effect size for each grade between the two genders for WRC ranged from small to medium. A few of the small sizes were close to the medium range.

Table 10

Effect Sizes for WRC – Females vs. Males 1996

	(Females)		(Males)		(Males)	
Grade	N	M	N	M	SD	Effect Size
One	165	39.93	153	33.14	27.38	0.25 small *
Two	159	82.60	155	64.21	34.12	0.54 medium *
Three	142	109.10	178	98.84	40.17	0.26 small *
Four	163	116.20	155	105.96	39.94	0.26 small *
Five	149	132.80	162	115.57	41.16	0.42 small *
Six	174	130.20	146	117.41	41.97	0.31 small *
Seven	164	145.90	155	128.60	37.86	0.46 small *

Words Read Correctly – Females vs. Males – 2003

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 11 for females versus males in Words Read Correctly (WRC) in 2003. The largest sample size for females was 171 in Grade 7 (see column 4, row 9) and the smallest sample size was 115 in Grade 1 (see column 4, row 3). The largest sample size for males was 164 in Grade 7 (see column 8, row 9) and the smallest sample size was 132 in Grade 1 (see column 8, row 3). The mean score increased with each successive

grade level for both genders. The male median S.E. 3.13 multiplied by 1.96, provided the 95% confidence interval of ± 6.13 used in Figure 5.

Table 11

Descriptive Statistics for WRC – Females vs. Males – 2003

	(Females)				(Males)			
Grade	M	SD	N	SE	M	SD	N	SE
One	40.03	30.89	115	2.88	32.53	28.09	132	2.44
Two	89.70	44.02	118	4.05	74.07	39.71	147	3.28
Three	115.20	35.21	124	3.16	106.42	42.25	157	3.37
Four	122.00	39.05	156	3.13	118.53	37.58	153	3.04
Five	135.30	39.48	141	3.32	125.58	37.05	135	3.19
Six	144.10	41.44	149	3.39	131.99	34.03	163	2.67
Seven	148.60	39.88	171	3.05	139.03	40.02	164	3.13

As illustrated in Figure 5, with the exception of Grade 4, there is significant difference ($p < .05$) in the mean scores of WRC Females versus Males in all grades in 2003. As can be seen in Figure 5, with the exception of Grade 4, the confidence intervals do not overlap for any of the grades. This signifies a continual difference between the two genders.

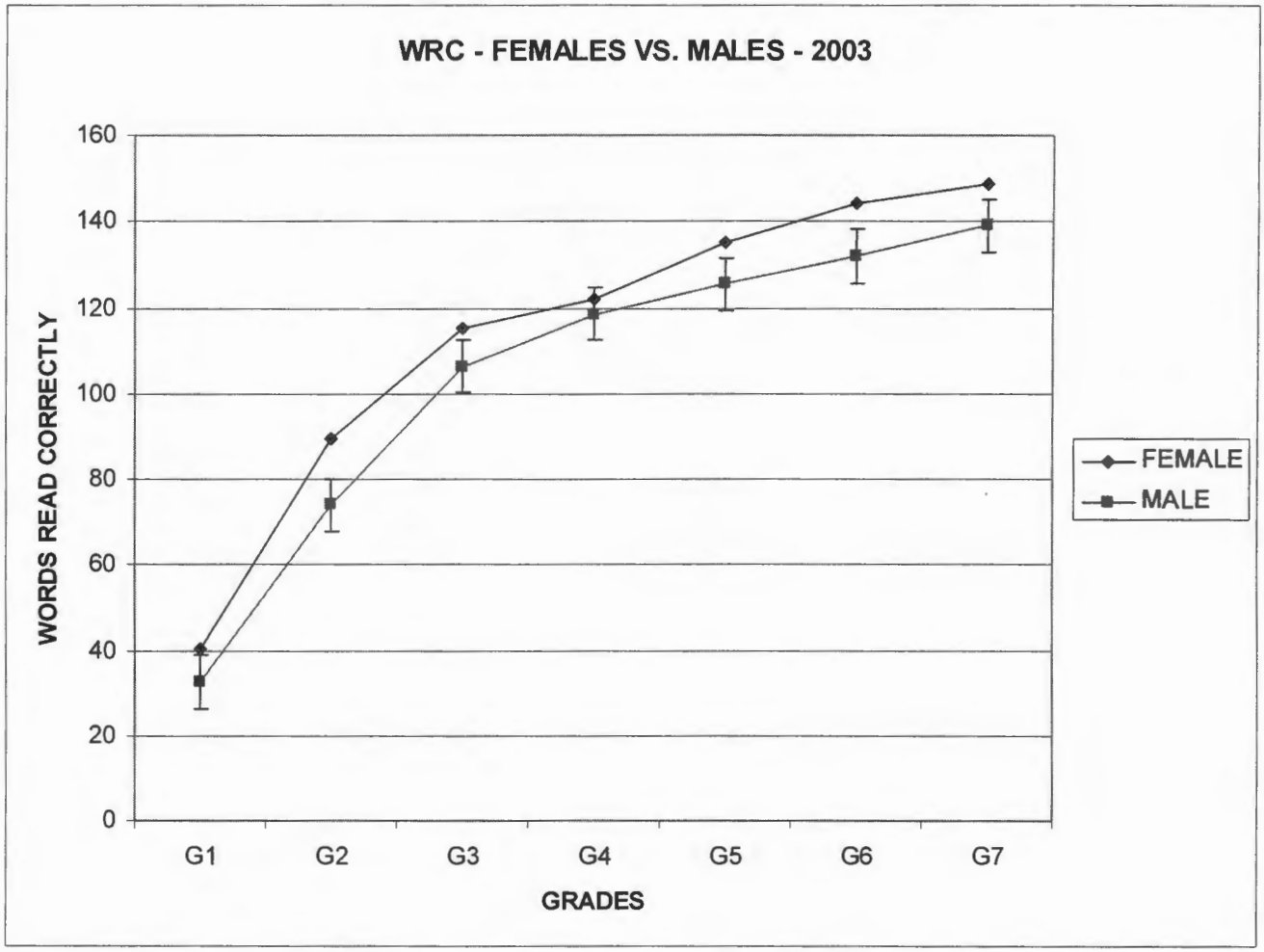

Figure 5. Grade Score Means for WRC – Females vs. Males – 2003

Cohen's *d* is calculated for WRC for each grade between each gender. The male standard deviation was used in the Cohen's *d* calculations. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 12, the effect size for each grade between the two genders for WRC was primarily small.

Table 12

Effect Sizes for WRC – Females vs. Males – 2003

	(Females)		(Males)		(Males)		
Grade	N	M	N	M	SD	Effect Size	
One	115	40.03	131	32.53	28.09	0.27	small *
Two	118	89.70	147	74.07	39.71	0.39	small *
Three	124	115.23	157	106.42	42.25	0.21	small *
Four	156	122.03	153	118.53	37.58	0.09	trivial
Five	141	135.34	135	125.58	37.07	0.26	small *
Six	149	144.11	163	131.99	34.03	0.36	small *
Seven	171	148.64	164	139.03	40.02	0.24	small *

Words Spelled Correctly – Females vs. Males – 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 13 for females versus males in Words Spelled Correctly (WSC) in 1996. The largest sample size for females was 173 in Grade 6 (see column 4, row 8) and the smallest sample size was 111 in Grade 2 (see column 4, row 4). The largest sample size for males was 160 in Grade 5 (see column 8, row 7) and the smallest sample size was 104 in Grade 2 (see column 8, row 4). The mean score increased with each successive grade level for both genders. The male median S.E. 1.09 multiplied by 1.96, provided the 95% confidence interval of ± 2.14 used in Figure 6.

Table 13

Descriptive Statistics for WSC – Females vs. Males – 1996

Grade	(Females) M	SD	N	SE	(Males) M	SD	N	SE
One	10.37	6.98	113	0.66	7.54	5.26	111	0.50
Two	26.18	10.72	111	1.02	19.45	8.73	104	0.86
Three	36.52	13.42	108	1.29	30.00	10.40	126	0.93
Four	44.73	15.40	162	1.21	35.76	13.60	156	1.09
Five	54.29	16.95	149	1.39	43.90	14.80	160	1.17
Six	60.12	14.41	173	1.09	48.52	14.30	144	1.19
Seven	65.77	16.58	164	1.29	55.43	17.00	156	1.36

As illustrated in Figure 6, there is significant difference ($p < .05$) in the mean scores of WSC Females versus Males in all grades in 1996. As can be seen in Figure 6, the confidence intervals do not overlap for any of the grades. This signifies a definite difference between the two genders.

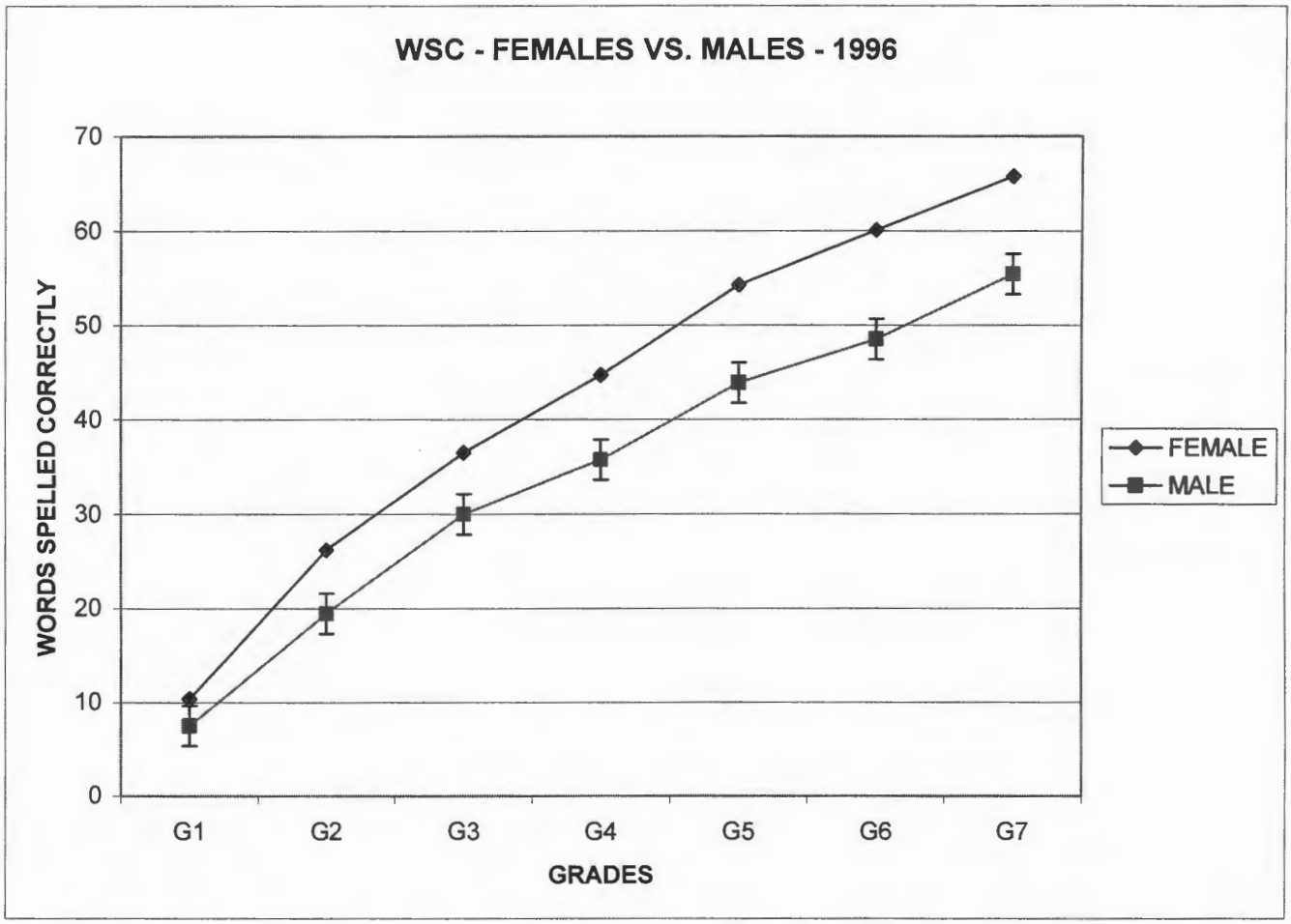

Figure 6. Grade Score Means for WSC – Females vs. Males – 1996

Cohen's *d* is calculated for WSC for each grade between each gender. The male standard deviation was used in the Cohen's *d* calculations. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 14, the effect size for each grade between the two genders for WSC was medium to large.

Table 14

Effect Sizes for WSC – Females vs. Males – 1996

	(Females)		(Males)		(Males)		
Grade	N	M	N	M	SD	Effect Size	
One	113	10.37	111	7.54	5.26	0.54	medium *
Two	111	26.18	104	19.45	8.73	0.77	medium *
Three	108	36.52	126	30.00	10.44	0.62	medium *
Four	162	44.73	156	35.76	13.64	0.66	medium *
Five	149	54.29	160	43.90	14.84	0.70	medium *
Six	173	60.12	144	48.52	14.29	0.81	large *
Seven	164	65.77	156	55.43	17.02	0.61	medium *

Words Spelled Correctly – Females vs. Males – 2003

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 15 for females versus males in Words Spelled Correctly (WSC) in 2003. The largest sample size for females was 170 in Grade 7 (see column 4, row 9) and the smallest sample size was 116 in Grade 1 (see column 4, row 4). The largest sample size for males was 164 in Grade 7 (see column 8, row 9) and the smallest sample size was 131 in Grade 1 (see column 8, row 3). The mean score increased with each successive grade level for both genders. The male median S.E. 1.16 multiplied by 1.96, provided the 95% confidence interval of ± 2.28 used in Figure 7.

Table 15

Descriptive Statistics for WSC – Females vs. Males – 2003

Grade	(Females) M	SD	N	SE	(Males) M	SD	N	SE
One	11.45	7.45	116	0.69	8.27	6.31	131	0.55
Two	25.08	12.32	118	1.13	20.67	9.03	147	0.75
Three	35.68	12.91	124	1.16	28.55	10.70	155	0.86
Four	46.38	14.53	156	1.16	39.79	14.50	153	1.17
Five	52.23	15.76	140	1.33	46.05	15.20	137	1.29
Six	61 84	17.66	149	1.45	52.48	15.80	162	1.24
Seven	68.86	16.91	170	1.30	57.51	14.90	164	1.16

As illustrated in Figure 7, there is significant difference ($p < .05$) in the mean scores of WSC Females versus Males in all grades in 2003. As can be seen in Figure 7, the confidence intervals do not overlap for any of the grades. This signifies a continuing difference between the two genders.

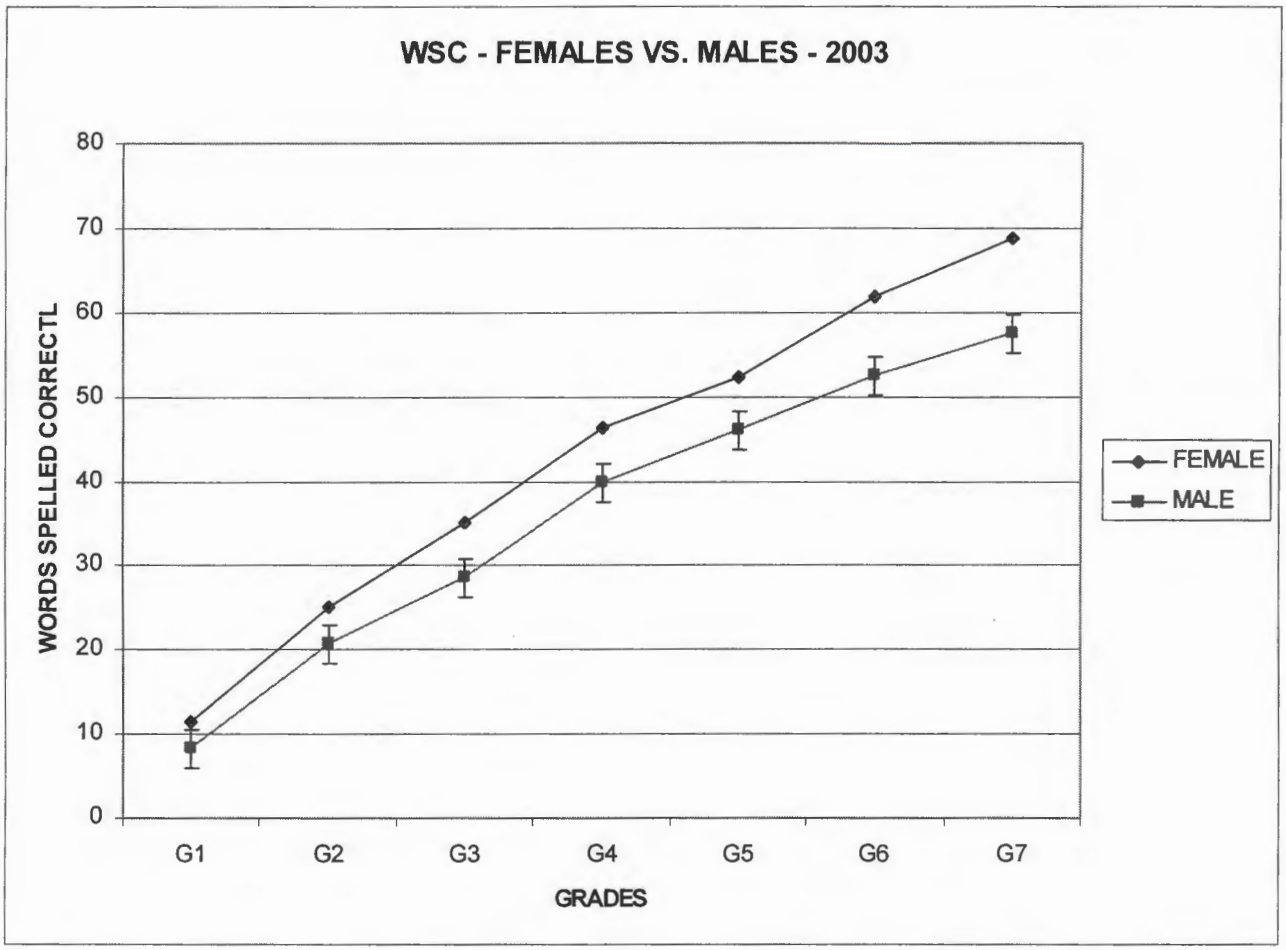

Figure 7. Grade Score Means for WSC – Females vs. Males – 2003

Cohen's *d* is calculated for WSC for each grade between each gender. The male standard deviation was used in the Cohen's *d* calculations. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 16, the effect size for each grade between the two genders for WSC ranged from small to medium.

Table 16

Effect Sizes for WSC – Females vs. Males – 2003

	(Females)		(Males)		(Males)		
Grade	N	M	N	M	SD	Effect Size	
One	116	11.45	131	8.27	6.31	0.50	medium *
Two	118	25.08	147	20.67	9.03	0.49	small *
Three	124	35.68	155	28.55	10.72	0.66	medium *
Four	156	46.38	153	39.79	14.47	0.46	small *
Five	140	52.23	137	46.05	15.18	0.41	small *
Six	149	61.84	162	52.48	15.79	0.59	medium *
Seven	170	68.86	164	57.51	14.87	0.76	medium *

Total Words Written – Females vs. Males – 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 17 for Females versus Males in Total Words Written (TWW) in 1996. The largest sample size for females was 173 in Grade 6 (see column 4, row 8) and the smallest sample size was 141 in Grade 3 (see column 4, row 5). The largest sample size for males was 178 in Grade 3 (see column 8, row 5) and the smallest sample size was 144 in Grade 6 (see column 8, row 8). The mean score increased with each successive grade level for both genders. The male median S.E. 1.11 multiplied by 1.96, provided the 95% confidence interval of ± 2.18 used in Figure 8.

Table 17

Descriptive Statistics for TWW – Females vs. Males – 1996

Grade	(Females) M	SD	N	SE	(Males) M	SD	N	SE
One	14.16	7.94	165	0.62	12.07	6.81	153	0.55
Two	29.32	10.89	157	0.87	23.01	8.88	155	0.71
Three	39.95	13.03	141	1.10	33.60	10.25	178	0.77
Four	48.10	15.15	162	1.19	38.84	13.81	156	1.11
Five	57.64	16.60	151	1.35	47.87	14.25	163	1.12
Six	62.40	14.57	173	1.11	51.79	14.57	144	1.21
Seven	68.00	16.51	164	1.29	57.94	17.31	156	1.39

As illustrated in Figure 8, there is significant difference ($p < .05$) in the mean scores of females versus males in all grades in 1996. As can be seen in Figure 8, the confidence intervals do not overlap for any of the grades. This signifies a definite difference between the two genders.

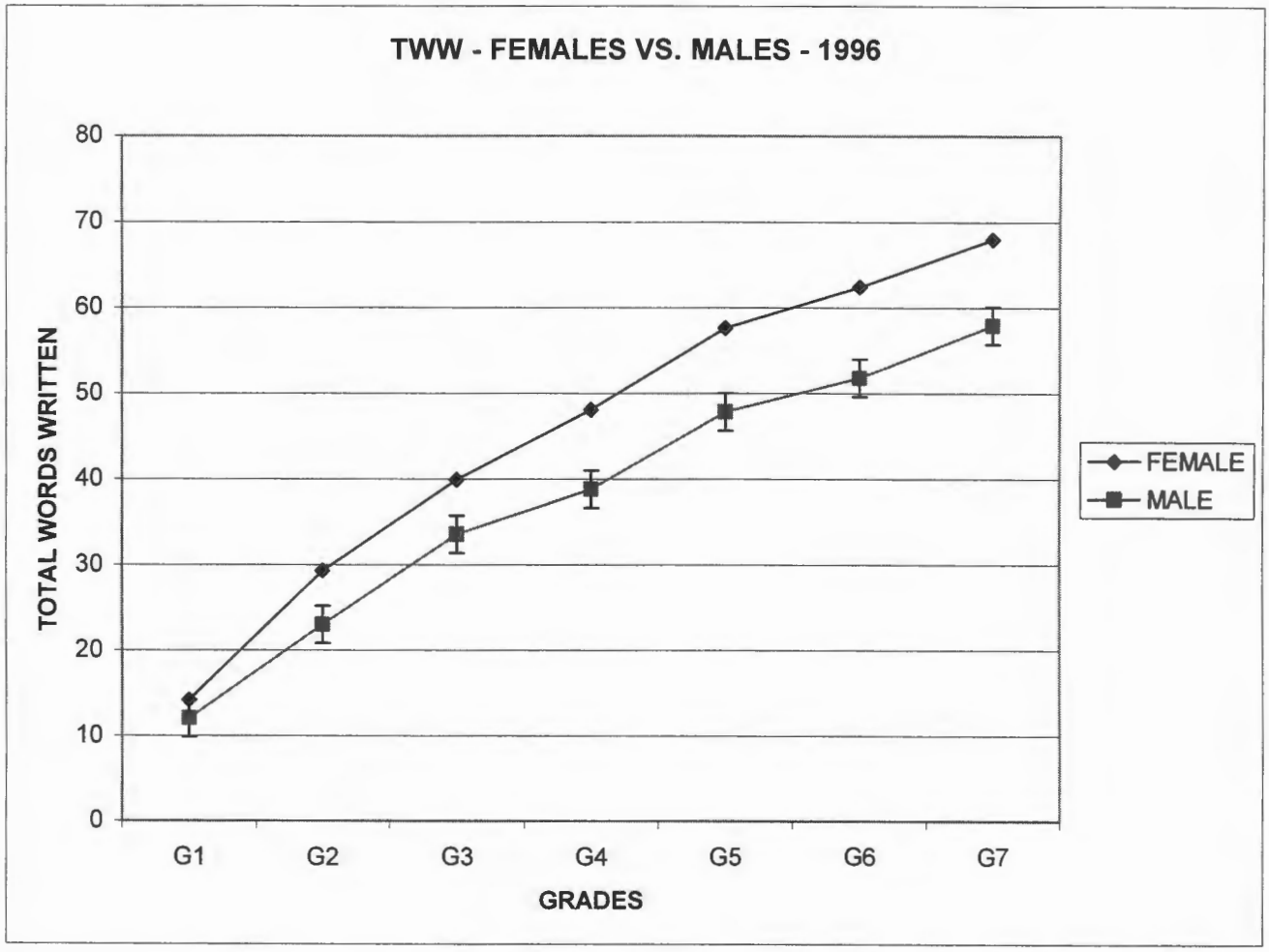

Figure 8. Grade Score Means for TWW – Females vs. Males – 1996

Cohen's *d* is calculated for TWW for each grade between each gender. The male standard deviation was used in the Cohen's *d* calculations. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 18, the effect size for each grade between the two genders for TWW ranged from small to high-medium.

Table 18

Effect Sizes for TWW – Females vs. Males – 1996

	(Females)		(Males)		(Males)		
Grade	N	M	N	M	SD	Effect	Size
One	165	14.16	153	12.07	6.81	0.31	small *
Two	157	29.32	155	23.01	8.88	0.71	medium *
Three	141	39.95	178	33.60	10.25	0.62	medium *
Four	162	48.10	156	38.84	13.81	0.67	medium *
Five	151	57.64	163	47.87	14.25	0.69	medium *
Six	173	62.40	144	51.79	14.57	0.73	medium *
Seven	164	68.00	156	57.94	17.31	0.58	medium *

Total Words Written – Females vs. Males – 2003

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 19 for Females versus Males in Total Words Written (TWW) in 2003. The largest sample size for females was 170 in Grade 7 (see column 4, row 9) and the smallest sample size was 116 in Grade 1 (see column 4, row 3). The largest sample size for males was 164 in Grade 7 (see column 8, row 9) and the smallest sample size was 131 in Grade 1 (see column 8, row 3). The mean score increased with each successive grade level for both genders. The male median S.E. 1.16 multiplied by 1.96, provided the 95% confidence interval of ± 2.27 used in Figure 9.

Table 19

Descriptive Statistics for TWW – Females vs. Males – 2003

	(Females)				(Males)			
Grade	M	SD	N	SE	M	SD	N	SE
One	15.28	8.45	116	0.78	11.83	7.79	131	0.68
Two	29.17	12.57	118	0.16	24.97	9.13	147	0.75
Three	38.76	13.16	124	1.18	32.02	10.89	155	0.87
Four	49.32	14.90	156	1.19	42.67	14.39	153	1.16
Five	54.36	15.60	140	1.32	48.85	15.51	137	1.32
Six	63.76	17.65	149	1.45	54.90	15.65	162	1.23
Seven	70.56	16.69	170	1.28	60.05	15.15	164	1.18

As illustrated in Figure 9, there is significant difference ($p < .05$) in the mean scores of TWW Females versus Males in all grades in 2003. As can be seen in Figure 9, the confidence intervals do not overlap for any of the grades. This signifies a continual difference between the two genders.

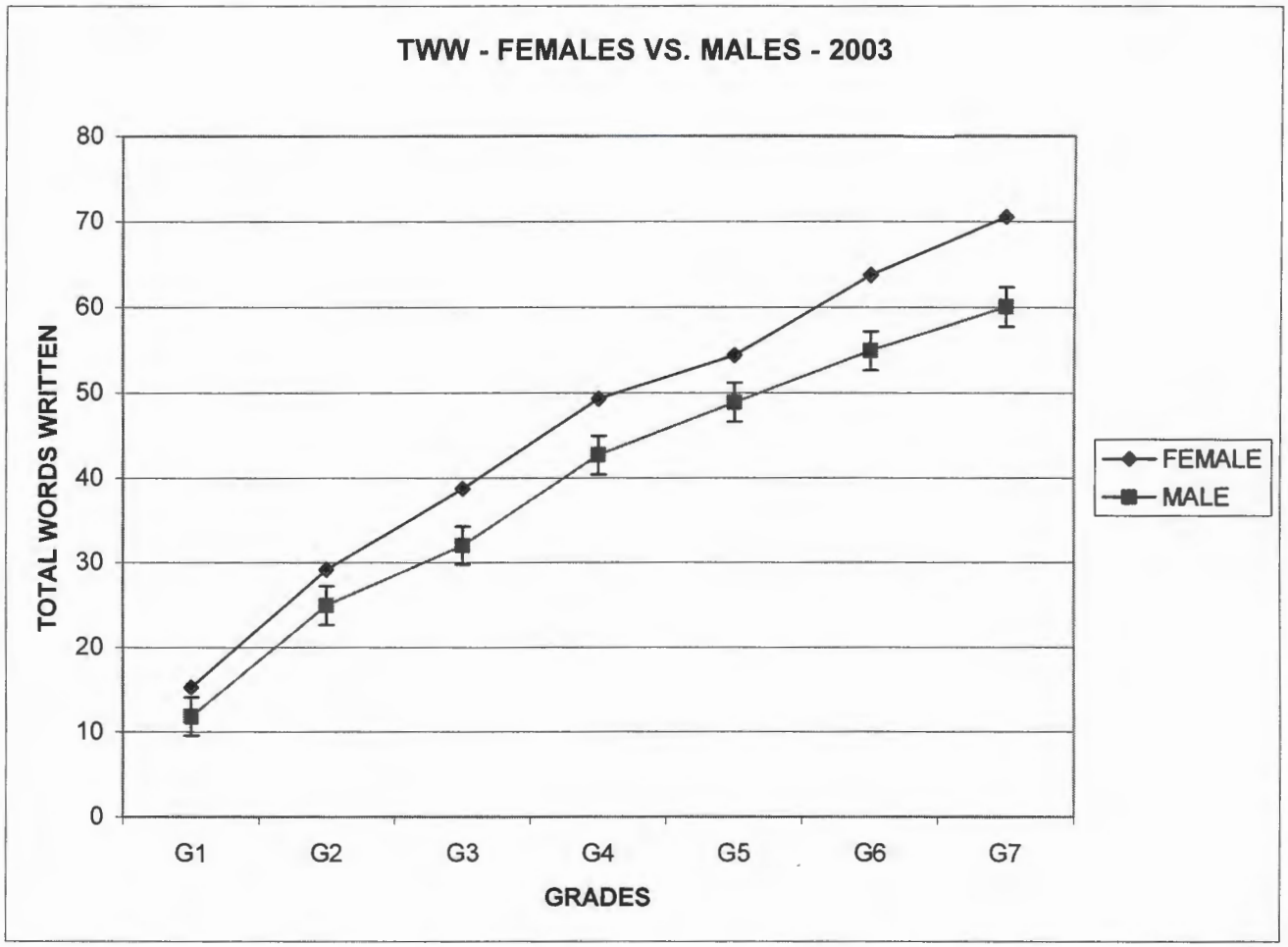

Figure 9. Grade Score Means for TWW – Females vs. Males – 2003

Cohen's *d* is calculated for TWW for each grade between each gender. The male standard deviation was used in the Cohen's *d* calculations. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 20, the effect size for each grade between the two genders for TWW ranged from small to medium.

Table 20

Effect Sizes for TWW – Females vs. Males – 2003

	(Females)		(Males)		(Males)	
Grade	N	M	N	M	SD	Effect Size
One	116	15.28	131	11.83	7.79	0.44 small *
Two	118	29.17	147	24.97	9.13	0.46 small *
Three	124	38.76	155	32.02	10.89	0.62 medium *
Four	156	49.32	153	42.67	14.39	0.46 small *
Five	140	54.36	137	48.85	15.51	0.36 small *
Six	149	63.76	162	54.90	15.65	0.57 medium *
Seven	170	70.56	164	60.05	15.15	0.69 medium *

Summary of Females versus Males 1996 and 2003

Words Read Correctly effect sizes show that the gap between females and males decreased in 2003 in five of seven grades (2, 3, 4, 5, 7), with Grade 7 having the greatest gap change of 0.28 (1996 effect size: 0.46; 2003 effect size: 0.24). Grade 1 and Grade 6 had a small increase in gap in 2003 (0.02 and 0.05).

Words Spelled Correctly effect sizes show that the gap between females and males decreased in 2003 in five of seven grades (1, 2, 4, 5, 6), with Grade 5 having the greatest gap change of 0.29 (1996 effect size: 0.70; 2003 effect size: 0.41). Grade 3 and Grade 7 had an increase in gap in 2003 (0.04 and 0.15).

Total Words Written effect sizes show that the gap between females and males decreased in 2003 in four of seven grades (2, 4, 5, 6), with Grade 5 having the greatest gap change of 0.33 (1996 effect size: 0.69; 2003 effect size: 0.36). Grade 1 and Grade 7 had an increase in gap in 2003 (0.13 and 0.11).

SERIES THREE: Introduction of Female 2003 versus Female 1996 and Male 2003 versus Male 1996

The decreasing gap between male and female performance in 2003 compared to 1996 could have resulted from a variety of sources (e.g. decrease in female scores with no decrease in male scores). Each gender was then compared in order to clarify the source of this changing gender gap. From this series of data we will be able to determine whether mean scores have improved in 2003 with each gender and which gender saw the greatest increase of mean scores.

Words Read Correctly – Females – 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 21 for females in Words Read Correctly (WRC) in 2003 and 1996. The largest sample size in 2003 was 171 in Grade 7 (see column 4, row 9) and the smallest sample size was 115 in Grade 1 (see column 4, row 3). The largest sample size in 1996 was 174 in Grade 6 (see column 8, row 8) and the smallest sample size was 142 in Grade 3 (see column 8, row 5). The mean score increased for each successive grade with both data sets, with the exception of 1996 where Grade 6 females scored lower than females in Grade 5. The 1996 median S.E. 3.30 multiplied by 1.96, provided the 95% confidence interval of ± 6.48 used in Figure 10.

Table 21

Descriptive Statistics for WRC – Females – 2003 vs. 1996

Grade	(2003) M	SD	N	SE	(1996) M	SD	N	SE
One	40.03	30.89	115	2.88	39.93	35.42	165	2.76
Two	89.70	44.02	118	4.05	82.60	34.45	159	2.73
Three	115.23	35.21	124	3.16	109.13	41.75	142	3.50
Four	122.03	39.05	156	3.13	116.22	43.03	163	3.37
Five	135.34	39.48	141	3.32	132.84	40.88	149	3.35
Six	144.11	41.44	149	3.39	130.21	39.90	174	3.03
Seven	148.64	39.88	171	3.05	145.92	42.31	164	3.30

As illustrated in Figure 10, there is significant difference ($p < .05$) in the mean scores of WRC Female 2003 versus 1996 for Grades 2 and 6. Grades 2 and 3 were just inside the confidence interval range. Grades 5 and 7 mean scores also increased, while Grade 1 remained the same.

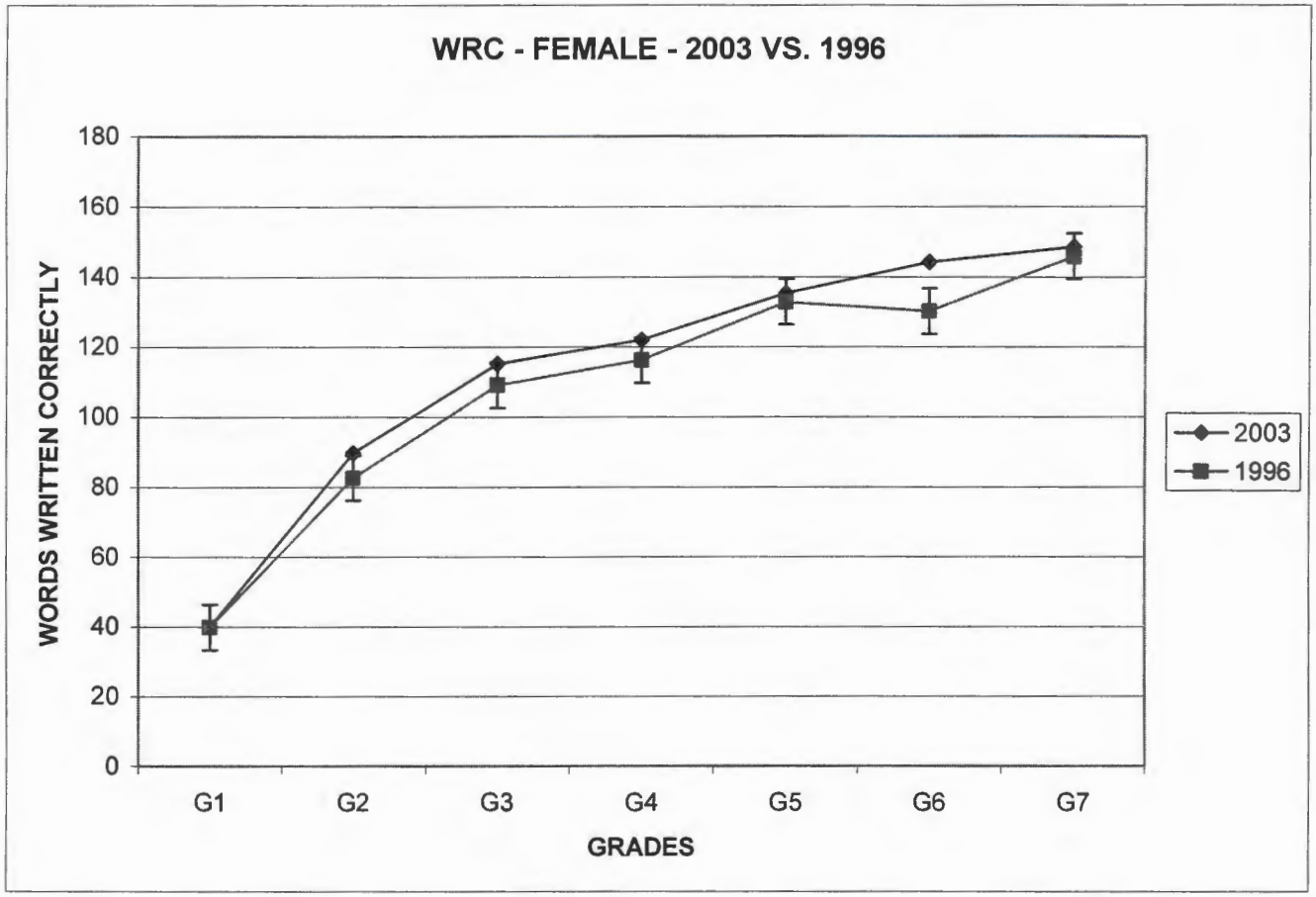

Figure 10. Grade Score Means for WRC – Females – 2003 vs. 1996

Cohen's *d* is calculated for WRC analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p <$ .05. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 22, the effect size for each grade between the two sets of data for WRC was either trivial or small.

Table 22

Effect Sizes for WRC – Females – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	115	40.03	165	39.93	35.42	0.00	trivial
Two	118	89.70	159	82.60	34.45	0.21	small *
Three	124	115.23	142	109.13	41.75	0.15	trivial
Four	156	122.03	163	116.22	43.03	0.14	trivial
Five	141	135.34	149	132.84	40.88	0.06	trivial
Six	149	144.11	174	130.21	39.90	0.35	small *
Seven	171	148.64	164	145.92	42.31	0.06	trivial

Words Read Correctly – Males – 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 23 for males in Words Read Correctly (WRC) in 2003 and 1996. The largest sample size in 2003 was 164 in Grade 7 (see column 4, row 9) and the smallest sample size was 132 in Grade 1 (see column 4, row 3). The largest sample size in 1996 was 178 in Grade 3 (see column 8, row 5) and the smallest sample size was 146 in Grade 6 (see column 8, row 8). The mean score increased for each successive grade with both data sets. The 1996 median S.E. 3.04 multiplied by 1.96, provided the 95% confidence interval of ± 5.96 used in Figure 11.

Table 23

Descriptive Statistics for WRC – Males – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	32.53	28.09	132	2.44	33.14	27.38	153	2.21
Two	74.07	39.71	147	3.27	64.21	34.12	155	2.74
Three	106.42	42.25	157	3.37	98.84	40.17	178	3.01
Four	118.53	37.58	153	3.04	105.96	39.94	155	3.21
Five	125.58	37.05	135	3.19	115.57	41.16	162	3.23
Six	131.99	34.03	163	2.67	117.41	41.96	146	3.47
Seven	139.03	40.02	164	3.13	128.60	37.86	155	3.04

As illustrated in Figure 11, there is significant difference ($p < .05$) in the mean scores of WRC Males 2003 versus 1996 for all grades except Grade 1. As can be seen in Figure 11, the confidence intervals do not overlap the scores for Grades 2 to 7, signifying a definite difference in scores between the two years for those grades.

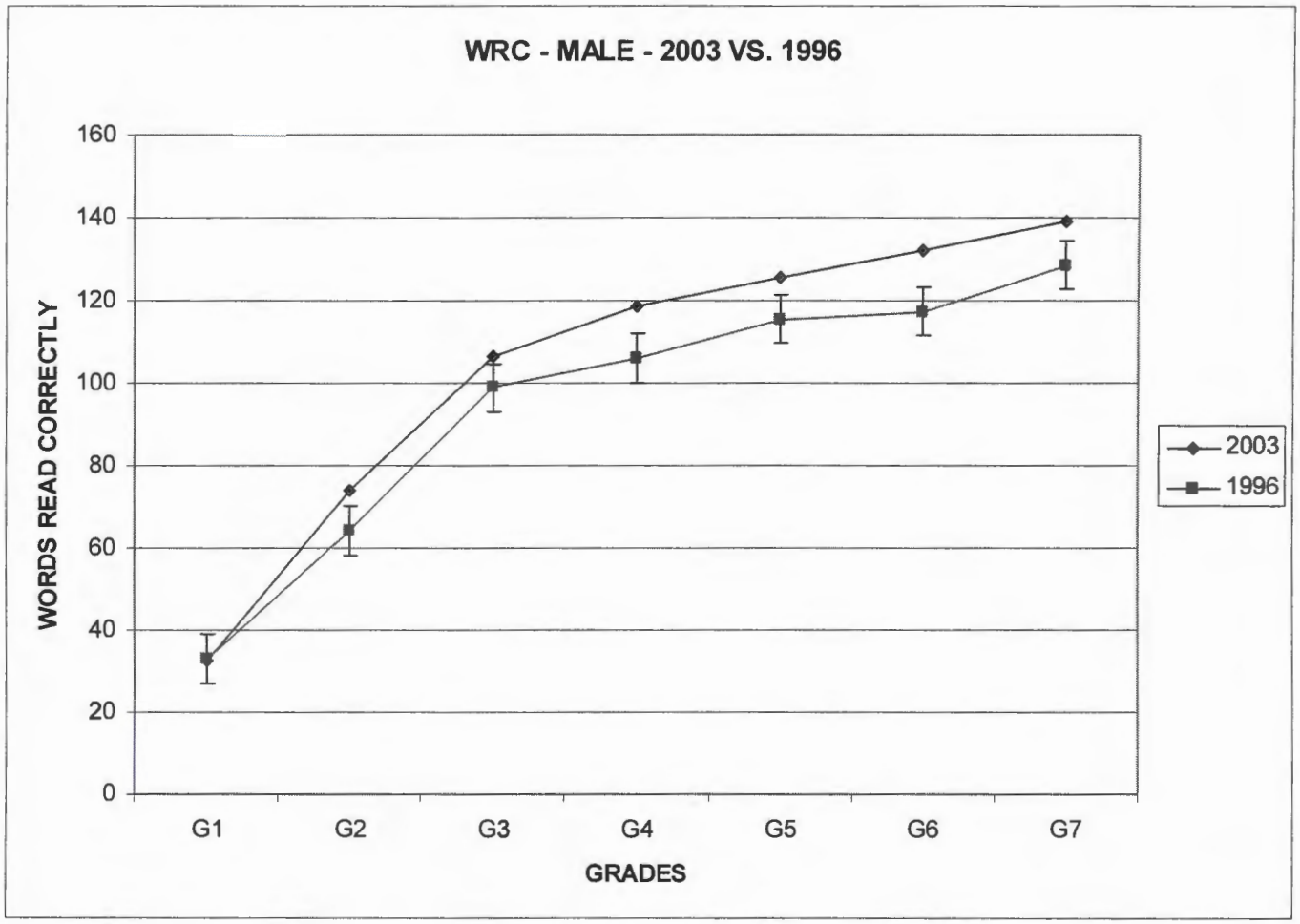

Figure 11. Grade Score Means for WRC – Males – 2003 vs. 1996

Cohen's *d* is calculated for WRC analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p <$ .05. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 24, the effect size for each grade between the two sets of data for WRC was primarily small.

Table 24

Effect Sizes for WRC – Males – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	132	32.53	153	33.14	27.38	0.02	trivial
Two	147	74.07	155	64.21	34.12	0.29	small *
Three	157	106.42	178	98.84	40.17	0.19	trivial *
Four	153	118.53	155	105.96	39.94	0.31	small *
Five	135	125.58	162	115.57	41.16	0.24	small *
Six	163	131.99	146	117.41	41.97	0.35	small *
Seven	164	139.03	155	128.60	37.86	0.28	small *

Words Spelled Correctly – Females – 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 25 for females in Words Spelled Correctly (WSC) in 2003 and 1996. The largest sample size in 2003 was 170 in Grade 7 (see column 4, row 9) and the smallest sample size was 116 in Grade 1 (see column 4, row 3). The largest sample size in 1996 was 173 in Grade 6 (see column 8, row 8) and the smallest sample size was 108 in Grade 3 (see column 8, row 5). The mean score increased for each successive grade in both data sets. The 1996 median S.E. 1.21 multiplied by 1.96, provided the 95% confidence interval of ± 2.37 used in Figure 12.

Table 25

Descriptive Statistics for WSC – Females – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	11.45	7.45	116	0.69	10.37	6.98	113	0.66
Two	25.08	12.32	118	1.13	26.18	10.70	111	1.02
Three	35.68	12.91	124	1.16	36.52	13.40	108	1.29
Four	46.38	14.53	156	1.16	44.73	15.40	162	1.21
Five	52.23	15.76	140	1.33	54.29	17.00	149	1.39
Six	61.84	17.66	149	1.45	60.12	14.40	173	1.10
Seven	68.86	16.91	170	1.30	65.77	16.60	164	1.29

As illustrated in Figure 12, there is significant difference ($p < .05$) in the mean scores of WSC Females 2003 versus 1996 for only Grade 7. As can be seen in Figure 12, the confidence intervals do not overlap the scores for Grade 7, signifying a definite difference in scores between the two years for that grade. Grade 1, 4, and 6 had an increase in scores, while the mean scores in Grades 2, 3, and 5 in 2003 regressed from the scores in 1996.

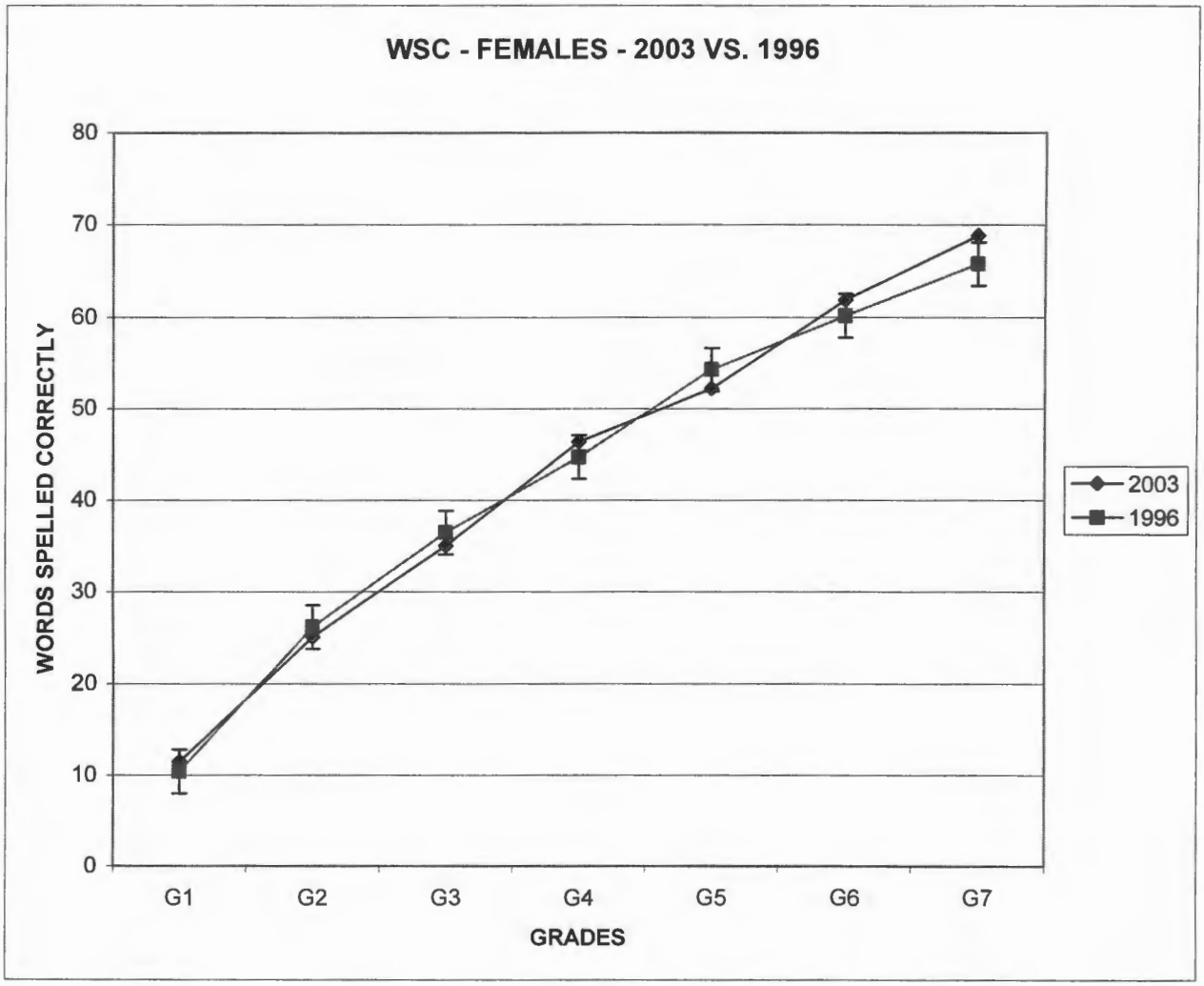

Figure 12. Grade Score Means for WSC – Females – 2003 vs. 1996

Cohen's *d* is calculated for WSC analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p <$ .05. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 26, the effect size for each grade between the two sets of data for WSC was trivial.

Table 26

Effect Sizes for WSC – Females – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	116	11.45	113	10.37	6.98	0.15	trivial
Two	118	25.08	111	26.18	10.72	- 0.10	trivial
Three	124	35.68	108	36.52	13.42	- 0.10	trivial
Four	156	46.38	162	44.73	15.40	0.11	trivial
Five	140	52.23	149	54.29	16.95	- 0.10	trivial
Six	149	61.84	173	60.12	14.41	0.12	trivial
Seven	170	68.86	164	65.77	16.58	0.19	trivial *

Words Spelled Correctly – Males – 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 27 for males in Words Spelled Correctly (WSC) in 2003 and 1996. The largest sample size in 2003 was 164 in Grade 7 (see column 4, row 9) and the smallest sample size was 131 in Grade 1 (see column 4, row 3). The largest sample size in 1996 was 160 in Grade 5 (see column 8, row 7) and the smallest sample size was 104 in Grade 2 (see column 8, row 4). The mean score increased for each successive grade with both data sets. The 1996 median S.E. 1.09 multiplied by 1.96, provided the 95% confidence interval of ± 2.15 used in Figure 13.

Table 27

Descriptive Statistics for WSC – Males – 2003 vs. 1996

Grade	(2003) M	SD	N	SE	(1996) M	SD	N	SE
One	8.27	6.31	131	0.55	7.54	5.26	111	0.50
Two	20.67	9.03	147	0.75	19.45	8.73	104	0.86
Three	28.55	10.72	155	0.86	30.00	10.44	126	0.93
Four	39.79	14.47	153	1.17	35.76	13.64	156	1.09
Five	46.05	15.16	137	1.29	43.90	14.84	160	1.17
Six	52.48	15.79	162	1.24	48.52	14.27	144	1.19
Seven	57.51	14.87	164	1.16	55.43	17.02	156	1.36

As illustrated in Figure 13, there is significant difference ($p < .05$) in the mean scores of WSC Males 2003 versus 1996 for Grades 4 and 6. As can be seen in Figure 13, the confidence intervals do not overlap the scores for Grades 4 and 6, signifying a definite difference in scores between the two years for those grades. Grades 5 and 7 scores finished just inside the confidence intervals.

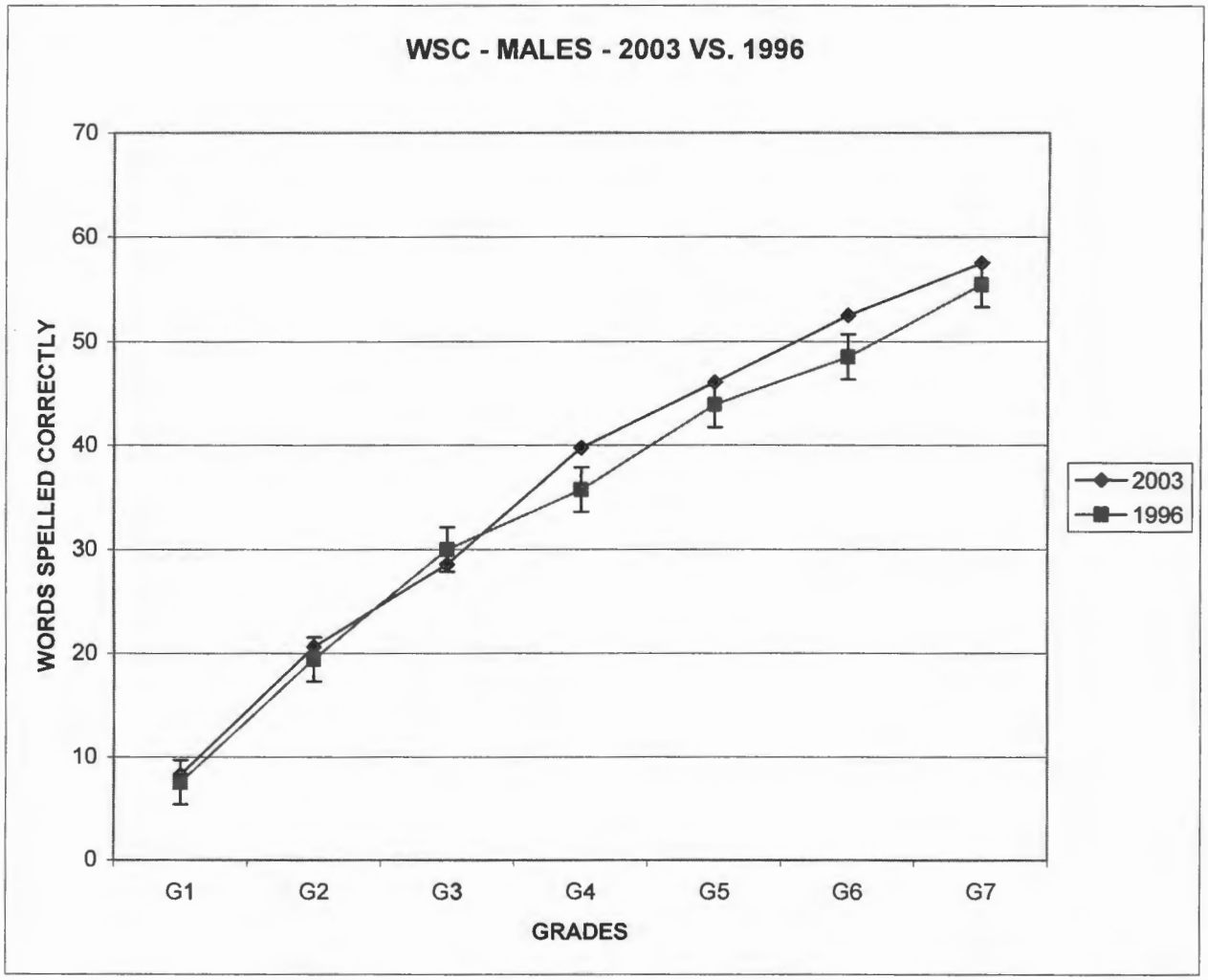

Figure 13. Grade Score Means for WSC – Males – 2003 vs. 1996

Cohen's *d* is calculated for WSC analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p <$ .05. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 28, the effect size for each grade between the two sets of data for WSC was either trivial or small.

Table 28

Effect Sizes for WSC – Males – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	131	8.27	111	7.54	5.26	0.14	trivial
Two	147	20.67	104	19.45	8.73	0.14	trivial
Three	155	28.55	126	30.00	10.44	- 0.14	trivial
Four	153	39.79	156	35.76	13.64	0.30	small *
Five	137	46.05	160	43.90	14.84	0.14	trivial
Six	162	52.48	144	48.52	14.29	0.28	small *
Seven	164	57.51	156	55.43	17.02	0.12	trivial

Total Words Written – Females – 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 29 for females in Total Words Written (TWW) in 2003 and 1996. The largest sample size in 2003 was 170 in Grade 7 (see column 4, row 9) and the smallest sample size was 116 in Grade 1 (see column 4, row 3). The largest sample size in 1996 was 173 in Grade 6 (see column 8, row 8) and the smallest sample size was 141 in Grade 3 (see column 8, row 5). The mean score increased for each successive grade in both data sets. The 1996 median S.E. 1.11 multiplied by 1.96, provided the 95% confidence interval of ± 2.18 used in Figure 14.

Table 29

Descriptive Statistics for TWW – Females – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	15.28	8.45	116	0.78	14.16	7.94	165	0.62
Two	29.17	12.57	118	1.16	29.32	10.89	157	0.87
Three	38.76	13.16	124	1.18	39.95	13.03	141	1.10
Four	49.32	14.90	156	1.19	48.10	15.16	162	1.19
Five	54.36	15.60	140	1.32	57.64	16.60	151	1.35
Six	63.76	17.65	149	1.45	62.40	14.57	173	1.11
Seven	70.56	16.69	170	1.28	68.07	16.51	164	1.29

As illustrated in Figure 14, there is significant difference ($p < .05$) in the mean scores of TWW Females 2003 versus 1996 for Grades 5 and 7. As can be seen in Figure 12, the confidence intervals do not overlap the scores for Grades 5 and 7, signifying a definite difference in scores between the two years for that grade. Of note is that the mean scores in Grades 3 and 5 in 2003 are lower than the mean scores for those grades in 1996, with Grade 5 significantly lower.

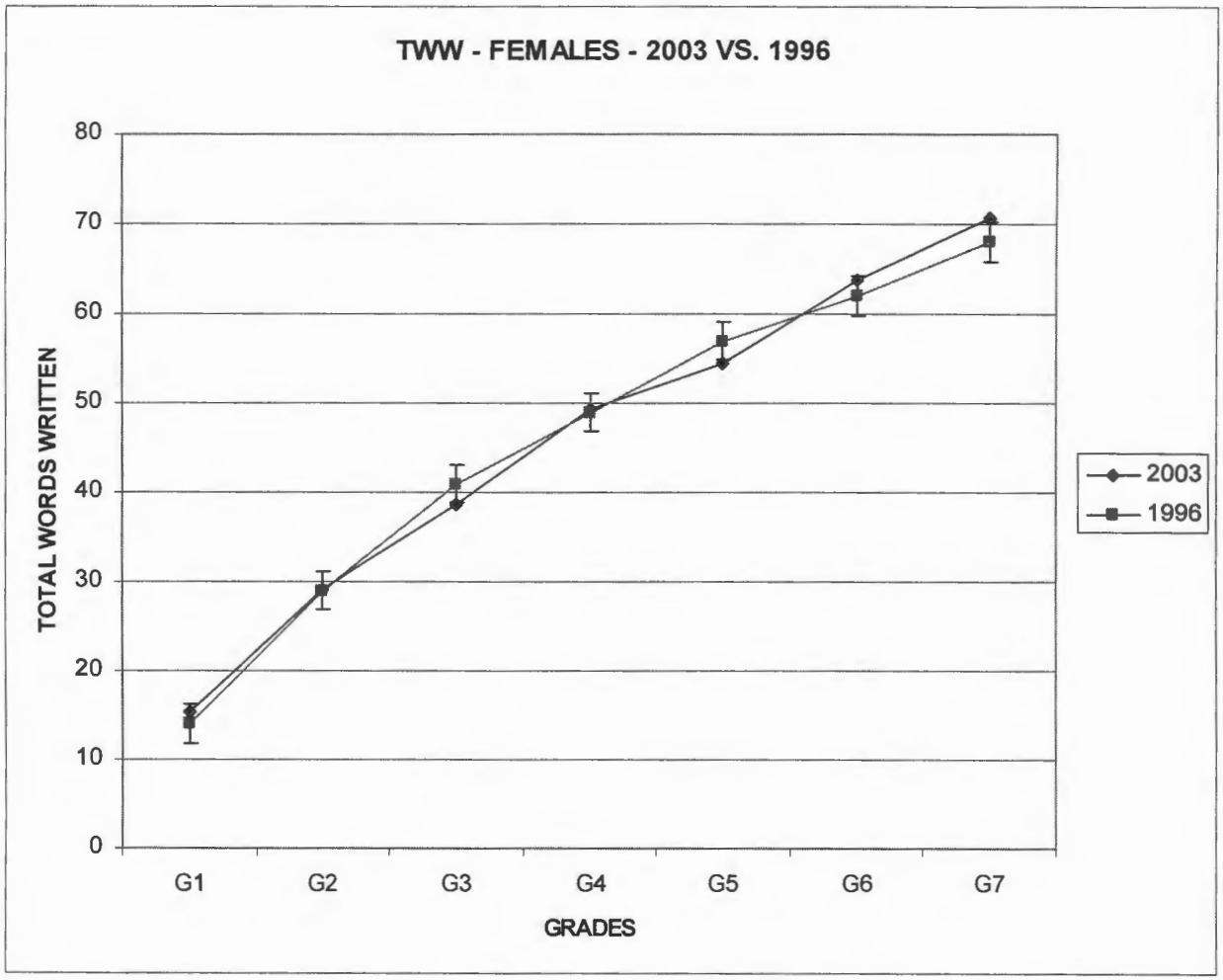

Figure 14. Grade Score Means for TWW – Females – 2003 vs. 1996

Cohen's *d* is calculated for TWW analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p <$ .05. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 30, the effect size for each grade between the two sets of data for TWW was primarily trivial.

Table 30

Effect Sizes for TWW – Females – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	116	15.28	165	14.16	7.94	0.14	trivial
Two	118	29.17	157	29.32	10.89	0.00	trivial
Three	124	38.76	141	39.95	13.03	- 0.10	trivial
Four	156	49.32	162	48.10	15.16	0.08	trivial
Five	140	54.36	151	57.64	16.60	- 0.20	small *
Six	149	63.76	173	62.40	14.57	0.09	trivial
Seven	170	70.56	164	68.07	16.51	0.15	trivial *

Total Words Written – Males – 2003 vs. 1996

The mean (M), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 31 for males in Total Words Written (TWW) in 2003 and 1996. The largest sample size in 2003 was 164 in Grade 7 (see column 4, row 9) and the smallest sample size was 131 in Grade 1 (see column 4, row 3). The largest sample size in 1996 was 178 in Grade 3 (see column 8, row 5) and the smallest sample size was 144 in grade 6 (see column 8, row 8). The mean score increased for each successive grade with both data sets. The 1996 median S.E. 1.11 multiplied by 1.96, provided the 95% confidence interval of ± 2.17 used in Figure 15.

Table 31

Descriptive Statistics for TWW – Males – 2003 vs. 1996

	(2003)				(1996)			
Grade	M	SD	N	SE	M	SD	N	SE
One	11.83	7.79	131	0.68	12.07	6.81	153	0.55
Two	24.97	9.13	147	0.75	23.01	8.88	155	0.71
Three	32.02	10.90	155	0.87	33.60	10.30	178	0.77
Four	42.67	14.40	153	1.16	38.84	13.80	156	1.11
Five	48.85	15.50	137	1.32	47.87	14.30	163	1.12
Six	54.90	15.70	162	1.23	51.79	14.60	144	1.21
Seven	60.05	15.10	164	1.18	57.94	17.30	156	1.39

As illustrated in Figure 15, there is significant difference ($p < .05$) in the mean scores of TWW Males 2003 versus 1996 for Grades 4 and 6. As can be seen in Figure 15, the confidence intervals do not overlap the scores for Grades 4 and 6, signifying a definite difference in scores between the two years for those grades. Of note is that there is a decrease in Grade 3 scores.

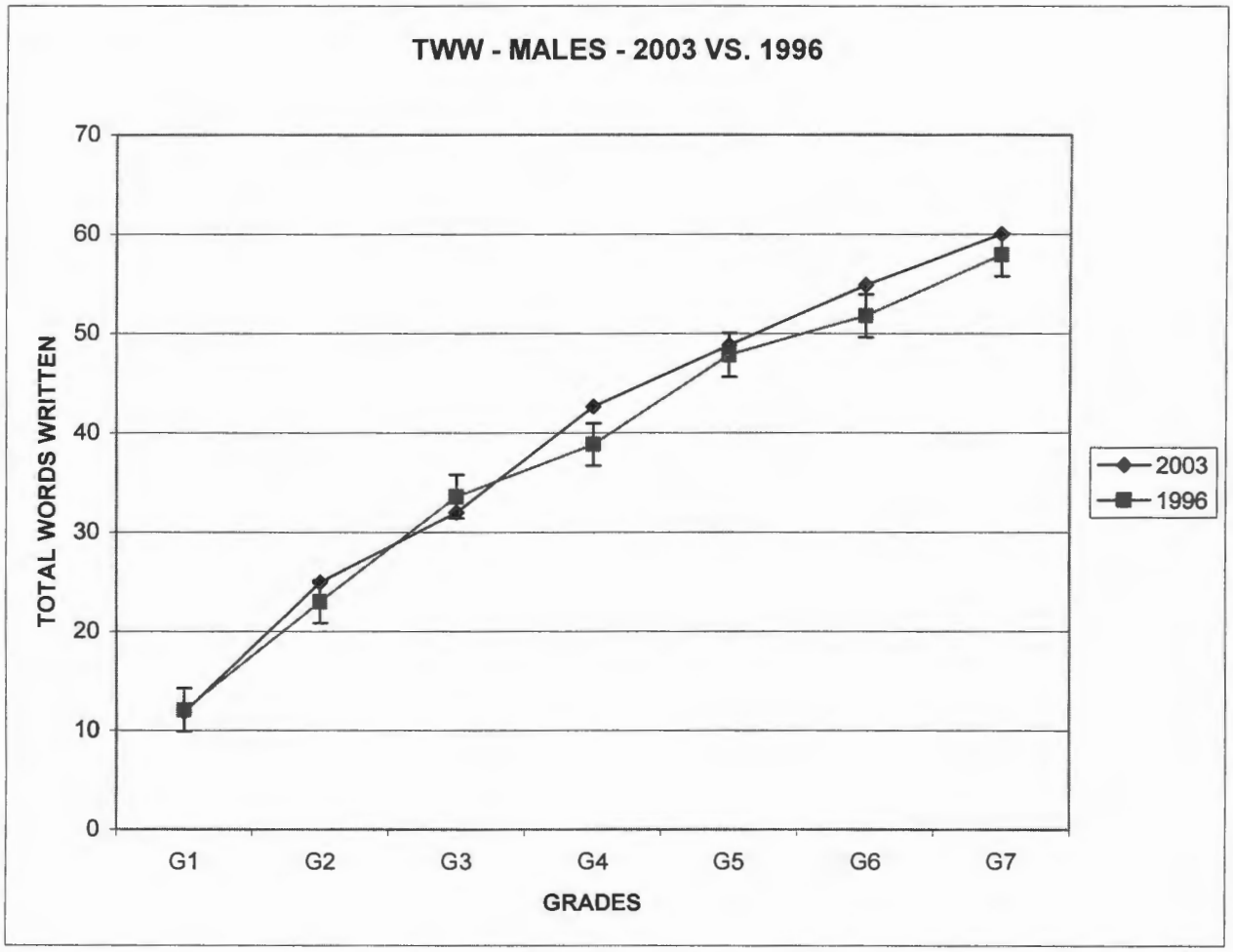

Figure 15. Grade Score Means for TWW – Males – 2003 vs. 1996

Cohen's *d* is calculated for TWW analysis for each grade for the 1996 - 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p <$ .05. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 32, the effect size for each grade between the two sets of data for TWW was either trivial or small.

Table 32

Effect Sizes for TWW – Males – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	M	N	M	SD	Effect Size	
One	131	11.83	153	12.07	6.81	0.00	trivial
Two	147	24.97	155	23.01	8.88	0.22	small
Three	155	32.02	178	33.60	10.25	- 0.15	trivial
Four	153	42.67	156	38.84	13.81	0.28	small *
Five	137	48.85	163	47.87	14.25	0.07	trivial
Six	162	54.90	144	51.79	14.57	0.21	small *
Seven	164	60.05	156	57.94	17.31	0.12	trivial

Summary of Female 2003 versus Female 1996 and Male 2003 versus Male 1996

Males clearly saw the greatest increase of mean scores in 2003 in comparison to females. In Words Read Correctly, females had an increase in mean scores in six grades, two significantly, while males also had an increase in mean scores in six grades, however all six were significant increases.

In Words Spelled Correctly, females had an increase in mean scores in four grades, one significantly, while males had an increase in six grades, two significantly. Of note is that females had a decrease in mean scores in 2003 in three grades, while males had a decrease in only one grade.

In Total Words Written, females had an increase in mean scores in three grades,

two significantly, while males had an increase in five grades, two of which were significant. Of note is that females had a decrease in mean scores in 2003 in two grades, while males had a decrease in only one grade.

SERIES FOUR: Introduction of 10th Percentile – 2003 versus 1996 (Combined Genders/Female vs. Female/Male vs. Male)

This series of statistics and analysis will look at the differences in 1996 versus 2003 of scores at the 10th percentile. The data will provide some insight as to which gender has made the greatest gains at the percentile that is sometimes used as a cut off for deciding student support.

Words Read Correctly – 10th Percentile – Combined Genders – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 33 for Words Read Correctly (WRC) for Combined Genders in 2003 and 1996. The researcher recognizes that the standard error is a standard deviation for the mean. The standard error of the median (50th percentile) is approximately 1.25 times as large as S.E. (Glass & Hopkins, 1996, p. 245). No standard error for another percentile (i.e. the 10th percentile) is described. For the remainder of this study the standard error of the median S.E. is used for illustrative purposes only. Therefore confidence intervals and adoptions of Cohen's *d* with the 10th percentile values cannot be strictly interpreted but merely serve to describe the degree of change. The 1996 median S.E. 2.30 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 5.64 used in Figure 16.

Table 33

Descriptive Statistics for WRC - 10th Percentile – Combined Genders – 2003 vs. 1996

	(2003)				(1996)			
Grade	P_{10}	SD	N	SE	P_{10}	SD	N	SE
One	7.00	29.60	247	1.88	6.00	31.94	318	1.79
Two	25.00	42.32	265	2.60	27.00	35.45	314	2.00
Three	62.00	39.47	281	2.35	49.10	41.13	320	2.30
Four	69.00	38.30	309	2.18	58.00	41.80	318	2.34
Five	80.70	38.55	276	2.32	68.20	41.86	311	2.37
Six	86.00	38.17	312	2.16	69.00	41.29	320	2.31
Seven	91.60	40.18	335	2.20	84.00	41.07	319	2.30

As illustrated in Figure 16, there is significant change ($p < .05$) in the scores of the two data sets for WRC at the 10th percentile for Combined Genders. With the exception of Grades 1 and 2, all grades in 2003 significantly increased scores in comparison to the 1996 scores. As can be seen in Figure 16, the confidence intervals do not overlap for Grades 3 to 7, signifying growth between the two years. Grade 1 and 2 scores at the 10th percentile were virtually unchanged.

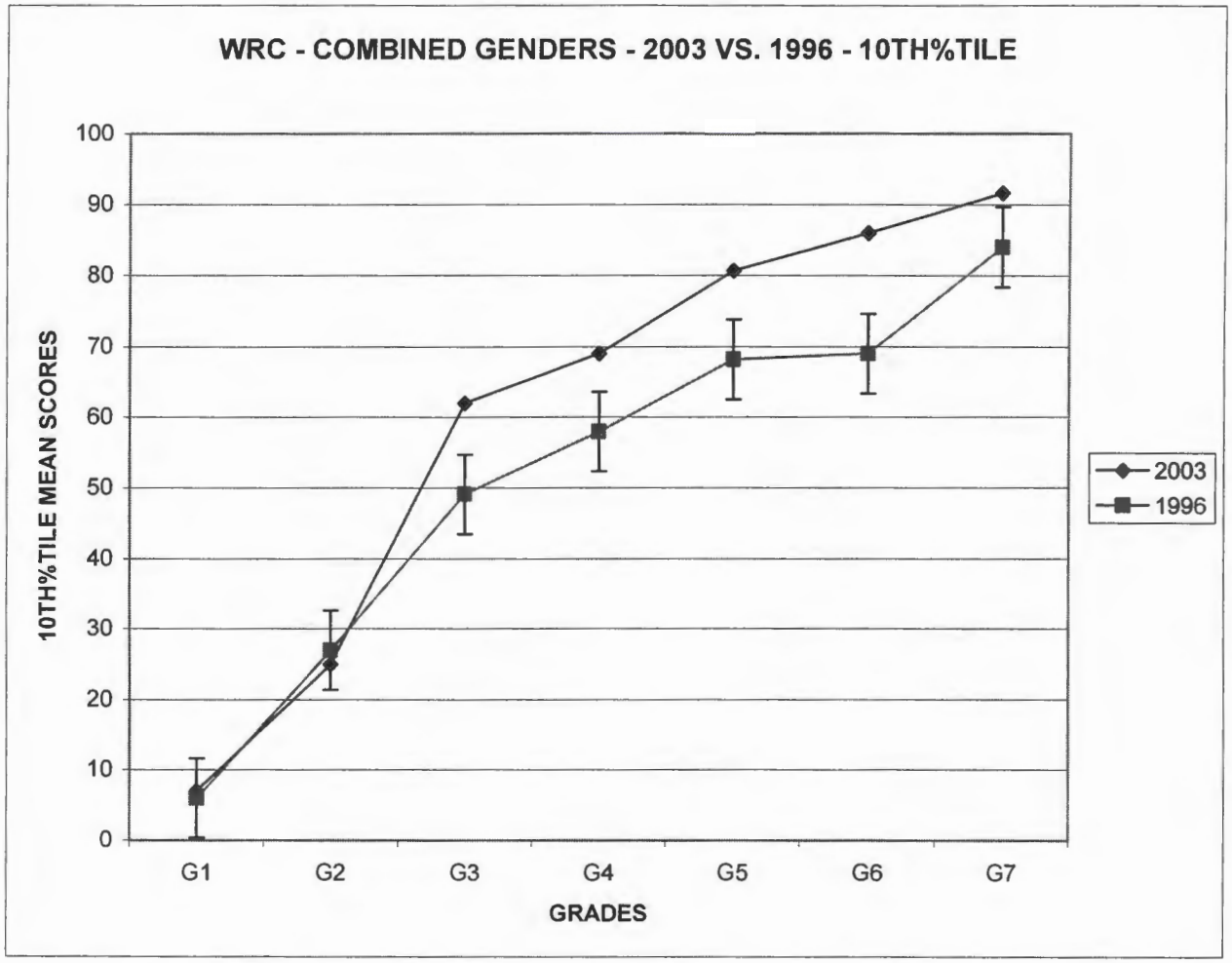

Figure 16. Grade Score Means in WRC – 10th Percentile – Combined Genders – 2003 vs. 1996.

Cohen's *d* is calculated for WRC analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes where there was a significant difference of $p < 0.05$ are marked with an asterisk. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 34, the effect size for each grade between the two data sets for WRC at the 10th percentile ranged from trivial to small.

Table 34

Effect Sizes for WRC – 10thPercentile – Combined Genders – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	P_{10}	N	P_{10}	SD	Effect Size	
One	247	7.00	318	6.00	31.94	0.03	trivial
Two	265	25.00	314	27.00	35.45	- 0.06	trivial
Three	281	62.00	320	49.10	41.13	0.31	small *
Four	309	69.00	318	58.00	41.80	0.26	small *
Five	276	80.70	311	68.20	41.86	0.30	small *
Six	312	86.00	320	69.00	41.29	0.41	small *
Seven	335	91.60	319	84.00	41.07	0.19	trivial *

Words Spelled Correctly – 10th Percentile – Combined Genders– 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 35 for Words Spelled Correctly (WSC) for Combined Genders in 2003 and 1996. The 1996 median S.E. 0.85 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 2.08 used in Figure 17 .

Table 35

Descriptive Statistics for WSC – 10th Percentile – Combined Genders – 2003 vs. 1996

	(2003)				(1996)			
Grade	P_{10}	SD	N	SE	P_{10}	SD	N	SE
One	2.00	7.03	247	0.45	2.00	6.33	224	0.42
Two	11.00	10.83	265	0.67	10.00	10.35	215	0.71
Three	17.00	12.25	279	0.73	18.50	12.32	234	0.81
Four	24.00	14.84	309	0.84	21.00	15.22	318	0.85
Five	30.80	15.74	277	0.95	29.00	16.70	309	0.95
Six	36.20	17.33	311	0.98	34.00	15.45	317	0.87
Seven	43.50	16.90	334	0.92	40.00	17.55	320	0.98

As illustrated in Figure 17, there is significant change ($p < .05$) in the scores of the two data sets for WSC at the 10th percentile for Combined Genders. Grades 4, 6, and 7 in 2003 significantly increased their scores in WSC at the 10th percentile in comparison to the 1996 scores. As can be seen in Figure 17, the confidence intervals do not overlap for Grades 4, 6, and 7, signifying growth at the 10th percentile from 1996 to 2003.

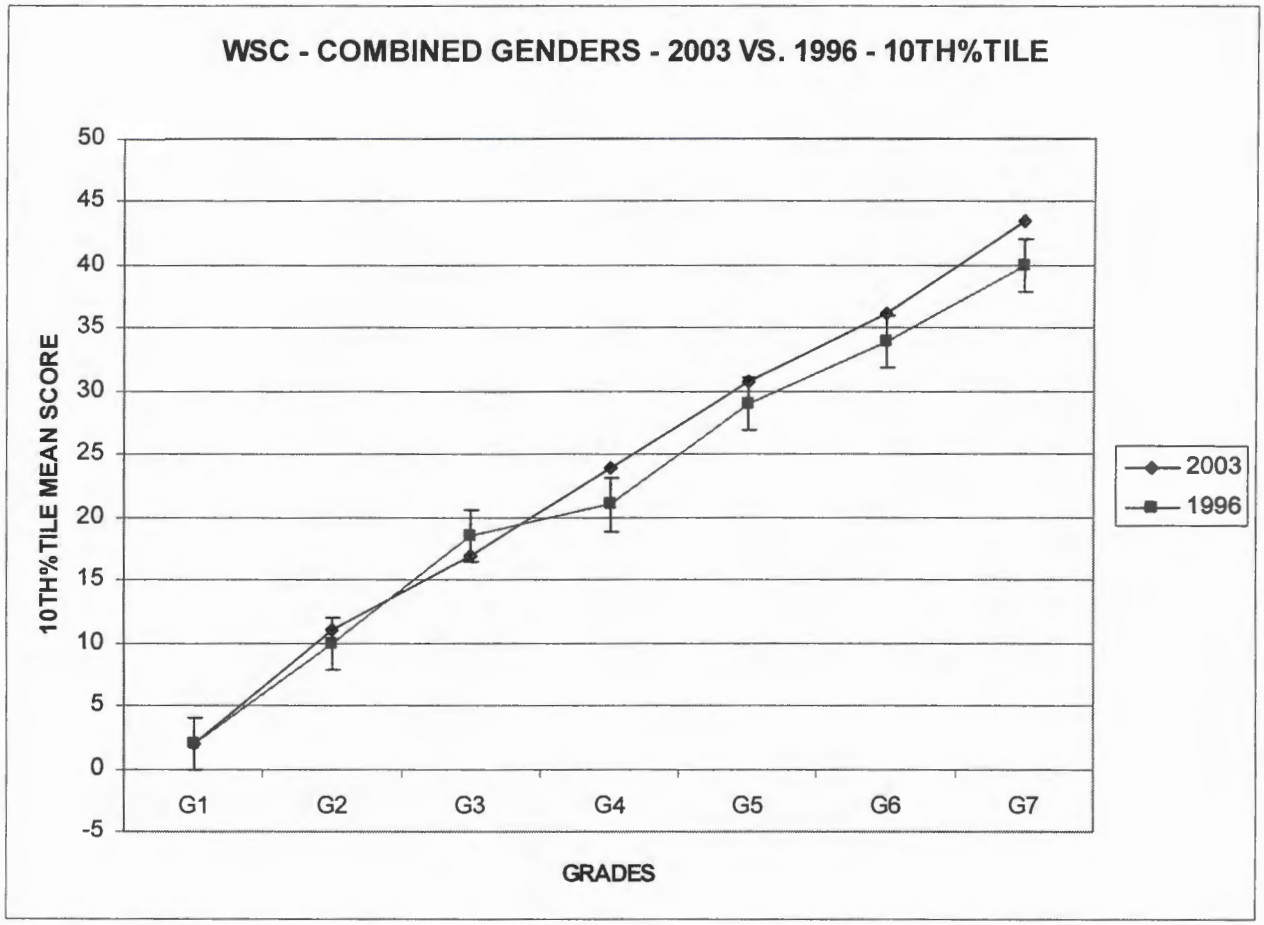

Figure 17. Grade Score Means in WSC – 10th Percentile – Combined Genders – 2003 vs. 1996

Cohen's *d* is calculated for WSC analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes where there was a significant difference of $p < .05$ are marked with an asterisk. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 36, the effect size for each grade between the two data sets for WSC at the 10th percentile was primarily trivial, although Grade 4 and 7 effect sizes nudged into the small range.

Table 36

Effect Sizes for WSC – 10th Percentile – Combined Genders – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	P_{10}	N	P_{10}	SD	Effect	Size
One	247	2.00	224	2.00	6.33	0.00	trivial
Two	265	11.00	215	10.00	10.35	0.10	trivial
Three	279	17.00	234	18.50	12.32	- 0.12	trivial
Four	309	24.00	318	21.00	15.22	0.20	small *
Five	277	30.80	309	29.00	16.70	0.11	trivial
Six	311	36.20	317	34.00	15.45	0.14	trivial *
Seven	334	43.50	320	40.00	17.55	0.20	small *

Total Words Written – 10th Percentile – Combined Genders– 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 37 for Total Words Written (TWW) for Combined Genders in 2003 and 1996. The 1996 median S.E. 0.92 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 2.25 used in Figure 18.

Table 37

Descriptive Statistics for TWW – 10th Percentile – Combined Genders – 2003 vs. 1996

	(2003)				(1996)			
Grade	P_{10}	SD	N	SE	P_{10}	SD	N	SE
One	4.00	8.28	247	0.53	4.00	7.48	318	0.42
Two	14.00	10.98	265	0.67	13.00	10.42	312	0.59
Three	21.00	12.39	279	0.74	21.00	11.97	319	0.67
Four	28.00	15.00	309	0.85	24.90	15.21	318	0.85
Five	34.00	15.77	277	0.95	33.00	16.16	314	0.91
Six	38.20	17.19	311	0.97	37.00	15.48	317	0.87
Seven	46.00	16.77	334	0.92	43.00	17.62	320	0.99

As illustrated in Figure 18, there is significant change ($p < .05$) in the scores of the two data sets for TWW at the 10th percentile for Combined Genders. Grades 4 and 7 in 2003 significantly increased their scores in TWW in comparison to the 1996 scores. As can be seen in Figure 18, the confidence intervals clearly do not overlap for Grades 4 and 7, signifying growth between the two data sets.

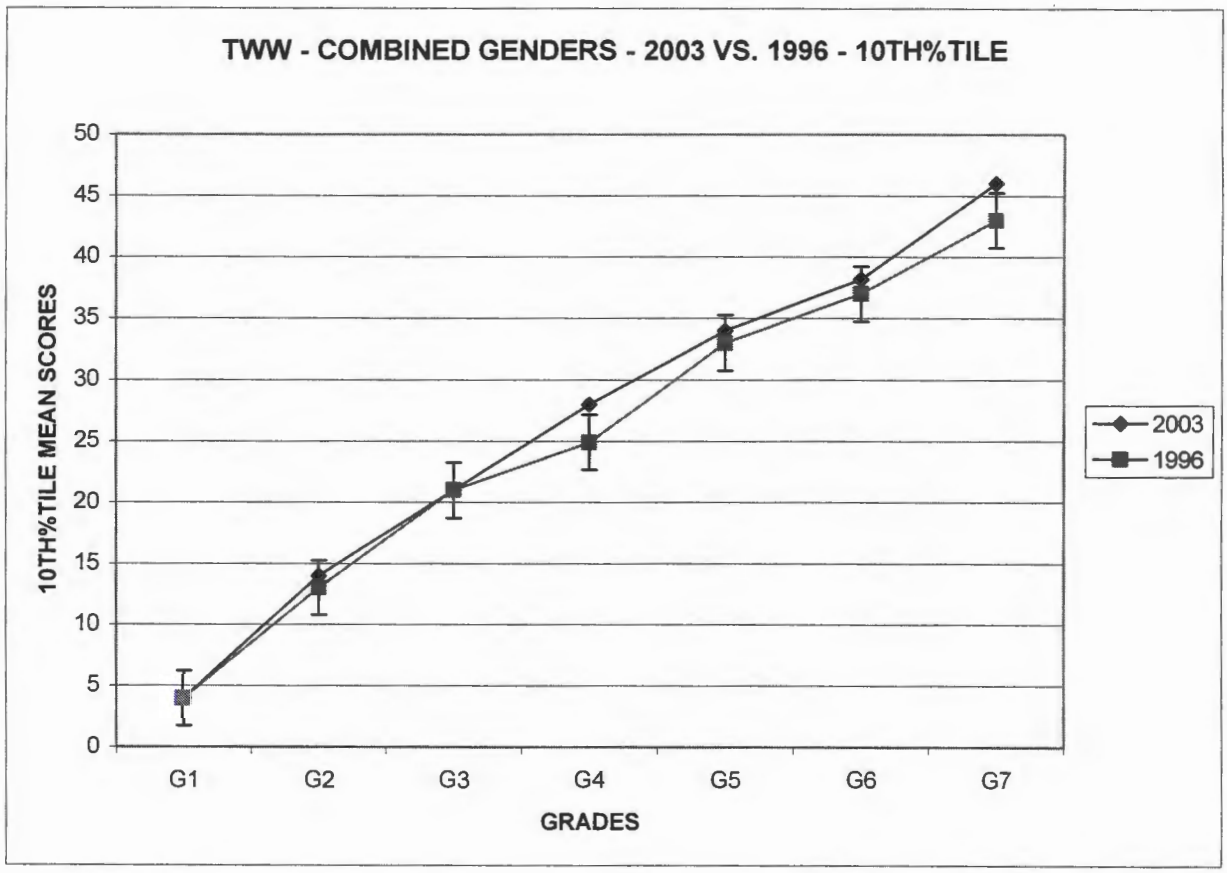

Figure 18. Grade Score Means in TWW – 10th Percentile – Combined Genders – 2003 vs. 1996

Cohen's *d* is calculated for TWW analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes where there was a significant difference of $p < .05$ are marked with an asterisk. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 38, the effect size for each grade between the two years for TWW at the 10th percentile was primarily trivial, although grade 7 was high trivial and grade 4 was small.

Table 38

Effect Sizes for TWW – 10th Percentile – Combined Genders – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	P_{10}	N	P_{10}	SD	Effect Size	
One	247	4.00	318	4.00	7.48	0.00	trivial
Two	265	14.00	312	13.00	10.42	0.10	trivial
Three	279	21.00	319	21.00	11.97	0.00	trivial
Four	309	28.00	318	24.90	15.21	0.20	small *
Five	277	34.00	314	33.00	16.16	0.06	trivial
Six	311	38.20	317	37.00	15.48	0.08	trivial
Seven	334	46.00	320	43.00	17.62	0.17	trivial *

Words Read Correctly – 10th Percentile – Females – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 39 for females in Words Read Correctly (WRC) at the 10th percentile in 2003 and 1996. The 1996 median S.E. 3.3 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 8.09 used in Figure 19.

Table 39

Descriptive Statistics for WRC – 10th Percentile – Females – 2003 vs. 1996

	(2003)				(1996)			
Grade	P_{10}	SD	N	SE	P_{10}	SD	N	SE
One	10.20	30.89	115	2.88	5.00	35.42	165	2.76
Two	31.60	44.02	118	4.05	35.00	34.45	159	2.73
Three	71.50	35.21	124	3.16	50.50	41.75	142	3.50
Four	74.00	39.05	156	3.13	57.80	43.03	163	3.37
Five	82.00	39.48	141	3.32	73.00	40.88	149	3.35
Six	86.00	41.44	149	3.39	73.50	39.90	174	3.03
Seven	95.20	39.88	171	3.05	91.50	42.31	164	3.30

As illustrated in Figure 19, there is significant difference ($p < .05$) in the scores of WRC at the 10th percentile for Females 2003 versus 1996 for Grades 3, 4, 5, and 6. As can be seen in Figure 19, the confidence intervals do not overlap the scores for Grades 3 to 6, signifying a definite difference in scores between the two years for those grades.

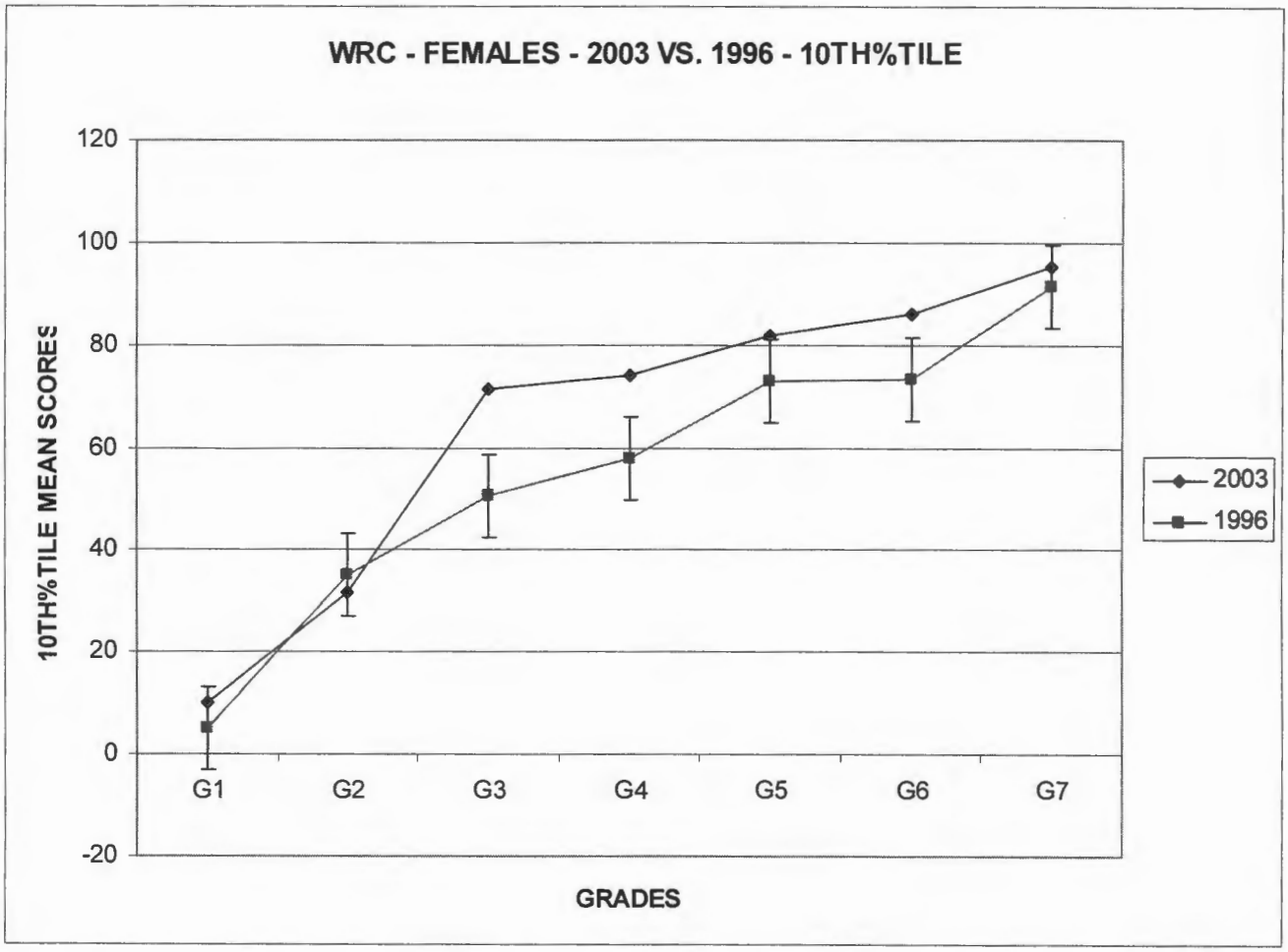

Figure 19. Grade Score Means for WRC – 10th Percentile – Females – 2003 vs. 1996

Cohen's *d* is calculated for WRC analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 40, the effect size for each grade between the two sets of data for WRC at the 10th percentile ranged from trivial to small to medium.

Table 40

Effect Sizes for WRC – 10th Percentile – Females – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	P_{10}	M	N	P_{10}	SD	Effect Size	
One	115	10.20	165	5.00	35.42	0.15	trivial
Two	118	31.60	159	35.00	34.45	- 0.10	trivial
Three	124	71.50	142	50.50	41.75	0.50	medium *
Four	156	74.00	163	57.80	43.03	0.38	small *
Five	141	82.00	149	73.00	40.88	0.22	small *
Six	149	86.00	174	73.50	39.90	0.31	small *
Seven	171	95.20	164	91.50	42.31	0.09	trivial

Words Read Correctly – 10th Percentile – Males – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 41 for males in Words Read Correctly (WRC) at the 10th percentile in 2003 and 1996. The 1996 median S.E. 3.04 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 7.45 used in Figure 20.

Table 41

Descriptive Statistics for WRC – 10th Percentile – Males – 2003 vs. 1996

Grade	(2003) P_{10}	SD	N	SE	(1996) P_{10}	SD	N	SE
One	5.30	28.09	132	2.44	6.00	27.38	153	2.21
Two	21.80	39.71	147	3.27	21.40	34.12	155	2.74
Three	47.40	42.25	157	3.37	49.00	40.17	178	3.01
Four	62.40	37.58	153	3.04	58.00	39.94	155	3.21
Five	77.60	37.05	135	3.19	62.00	41.16	162	3.23
Six	81.80	34.03	163	2.67	62.40	41.96	146	3.47
Seven	90.00	40.02	164	3.13	75.20	37.86	155	3.04

As illustrated in Figure 20, there is significant difference ($p < .05$) in the scores of WRC at the 10th percentile for Males 2003 versus 1996 for Grades 5, 6, and 7. As can be seen in Figure 20, the confidence intervals do not overlap the scores for Grades 5 to 7, signifying a definite difference in scores between the two years for the grades.

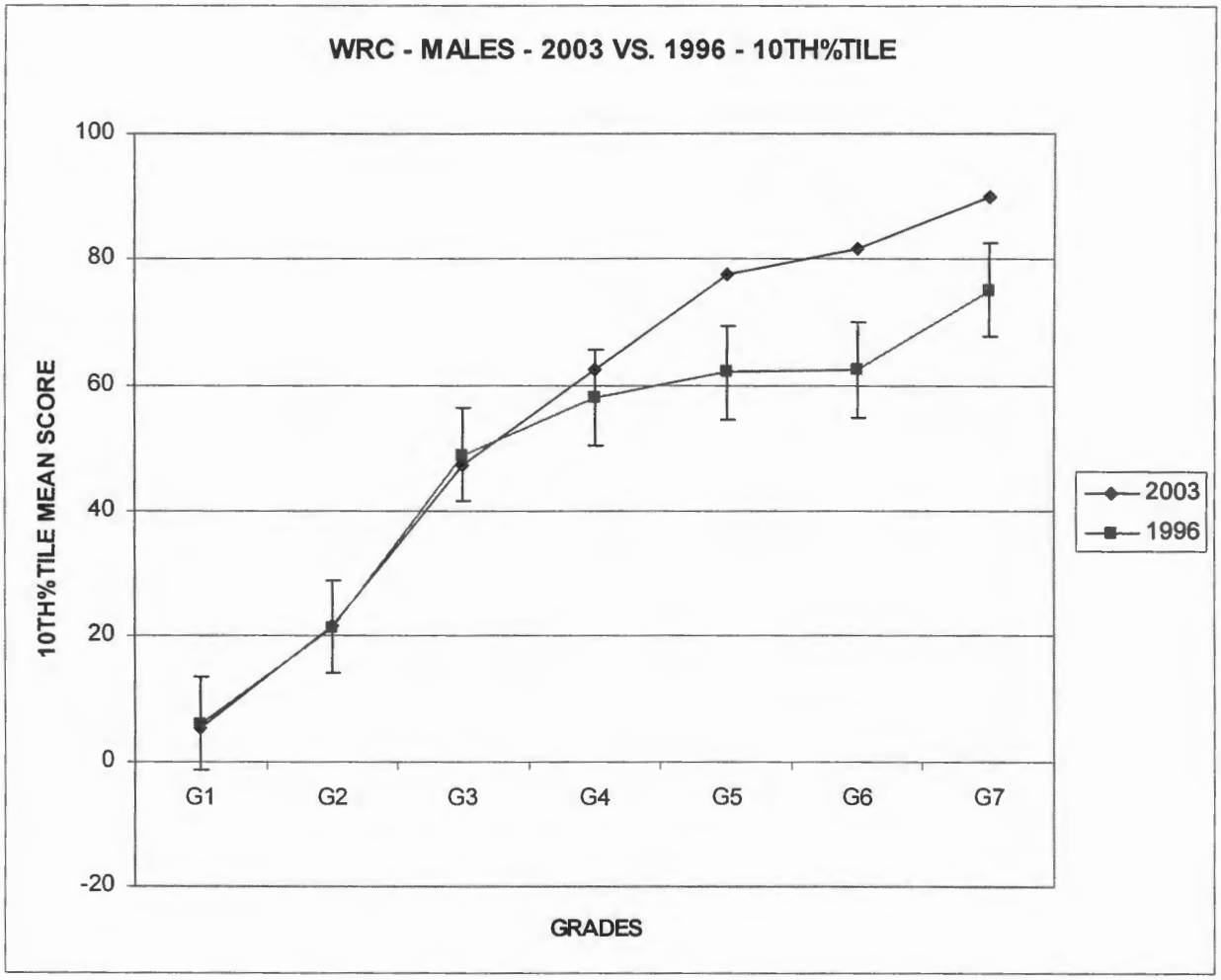

Figure 20. Grade Score Means for WRC – 10th Percentile – Males – 2003 vs. 1996

Cohen's *d* is calculated for WRC analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 42, the effect size for each grade between the two sets of data for WRC at the 10th percentile ranged from trivial in the lower grades to small in the upper grades.

Table 42

Effect Sizes for WRC – 10th Percentile – Males – 2003 vs. 1996

	(2003)		(1996)		(1996)	
Grade	N	P_{10}	N	P_{10}	SD	Effect Size
One	132	5.30	153	6.00	27.38	0.03 trivial
Two	147	21.80	155	21.40	34.12	0.01 trivial
Three	157	47.40	178	49.00	40.17	- 0.04 trivial
Four	153	62.40	155	58.00	39.94	0.11 trivial
Five	135	77.60	162	62.00	41.16	0.38 small *
Six	163	81.80	146	62.40	41.97	0.46 small *
Seven	164	90.00	155	75.20	37.86	0.39 small *

Words Spelled Correctly – 10th Percentile – Females – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 43 for females in Words Spelled Correctly (WSC) at the 10th percentile in 2003 and 1996. The 1996 median S.E. 1.21 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 2.97 used in Figure 21.

Table 43

Descriptive Statistics for WSC – 10th Percentile – Females – 2003 vs. 1996

	(2003)				(1996)			
Grade	P_{10}	SD	N	SE	P_{10}	SD	N	SE
One	3.00	7.45	116	0.69	2.00	6.98	113	0.66
Two	11.00	12.32	118	1.13	12.20	10.70	111	1.02
Three	20.00	12.91	124	1.16	19.90	13.40	108	1.29
Four	28.00	14.53	156	1.16	26.30	15.40	162	1.21
Five	33.10	15.76	140	1.33	37.00	17.00	149	1.39
Six	39.00	17.66	149	1.45	44.40	14.40	173	1.10
Seven	48.20	16.91	170	1.30	44.50	16.60	164	1.29

Scores of WSC at the 10th percentile for Females 2003 versus 1996 have some interesting results as seen in Figure 21.Grades 5, 6, and 7 show significant differences between the two data sets at the 10th percentile. While Grade 7 shows significant growth in 2003, Grades 5 and 6 show some significant regression.

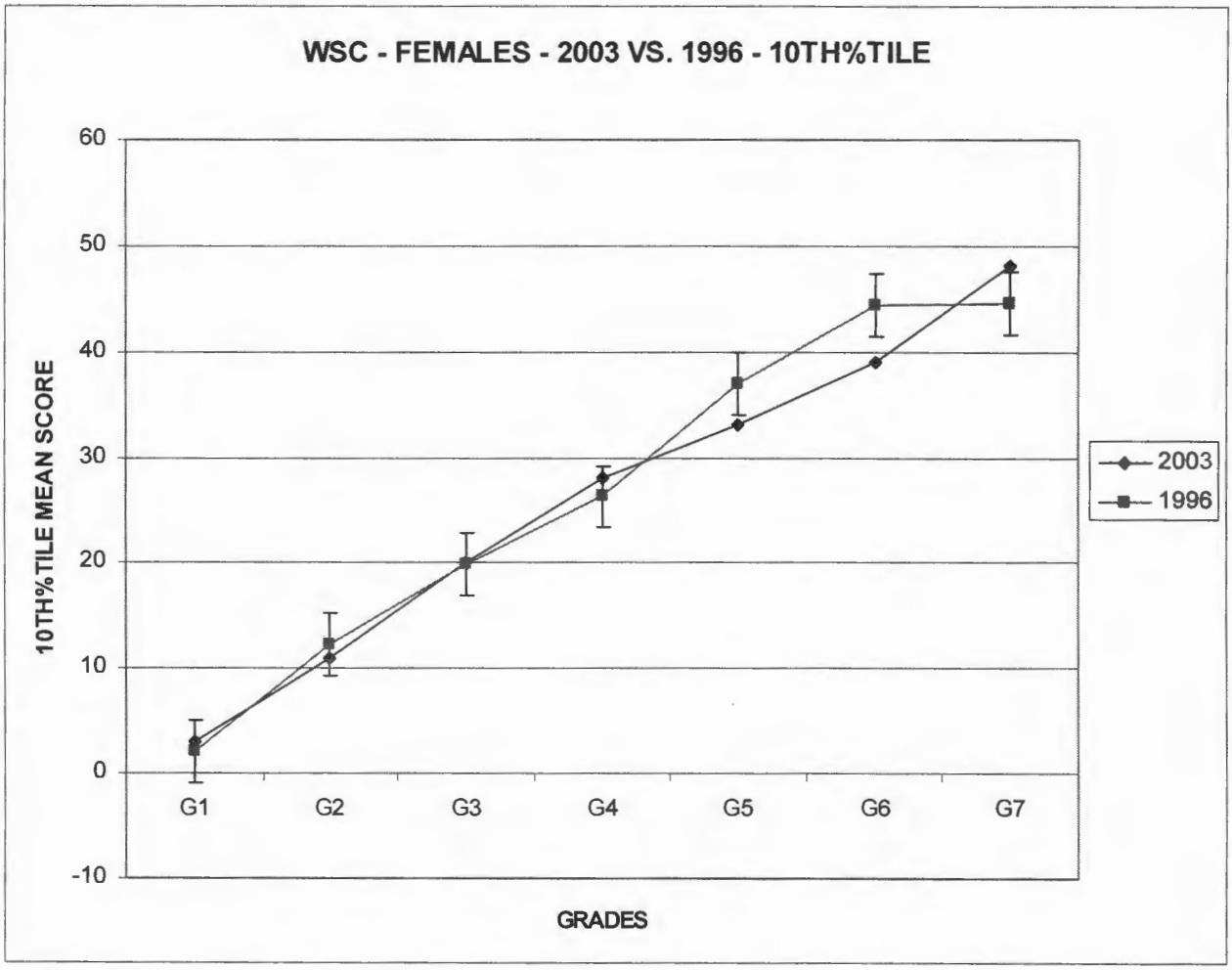

Figure 21. Grade Score Means for WSC – 10th Percentile – Females – 2003 vs. 1996

Cohen's *d* is calculated for WSC analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 44, the effect size for each grade between the two sets of data for WSC at the 10th percentile ranged from trivial to small.

Table 44

Effect Sizes for WSC – 10th Percentile – Females – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	P_{10}	N	P_{10}	SD	Effect Size	
One	116	3.00	113	2.00	6.98	0.14	trivial
Two	118	11.00	111	12.20	10.72	- 0.11	trivial
Three	124	20.00	108	19.90	13.42	0.01	trivial
Four	156	28.00	162	26.30	15.40	0.11	trivial
Five	140	33.10	149	37.00	16.95	- 0.23	small *
Six	149	39.00	173	44.40	14.41	- 0.37	small *
Seven	170	48.20	164	44.50	16.58	0.22	small *

Words Spelled Correctly – 10th Percentile – Males – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 45 for males in Words Spelled Correctly (WSC) at the 10th percentile in 2003 and 1996. The 1996 median S.E. 1.09 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 2.67 used in Figure 22.

Table 45

Descriptive Statistics for WSC – 10th Percentile – Males – 2003 vs. 1996

Grade	(2003) P_{10}	SD	N	SE	(1996) P_{10}	SD	N	SE
One	1.00	6.31	131	0.55	1.20	5.26	111	0.50
Two	10.80	9.03	147	0.75	9.00	8.73	104	0.86
Three	16.00	10.72	155	0.86	17.00	10.44	126	0.93
Four	22.40	14.47	153	1.17	20.00	13.64	156	1.09
Five	27.80	15.16	137	1.29	24.00	14.84	160	1.17
Six	34.00	15.79	162	1.24	31.00	14.27	144	1.19
Seven	39.00	14.87	164	1.16	36.70	17.02	156	1.36

As illustrated in Figure 22, there is significant difference ($p < .05$) in the scores of WSC at the 10th percentile for Males 2003 versus 1996 Grades 5 and 6. As can be seen in Figure 22, the confidence intervals do not overlap the scores for Grades 5 and 6, signifying a definite difference in scores between the two years for those grades. Grades 4 and 7 improved scores for 2003 were just inside the confidence intervals.

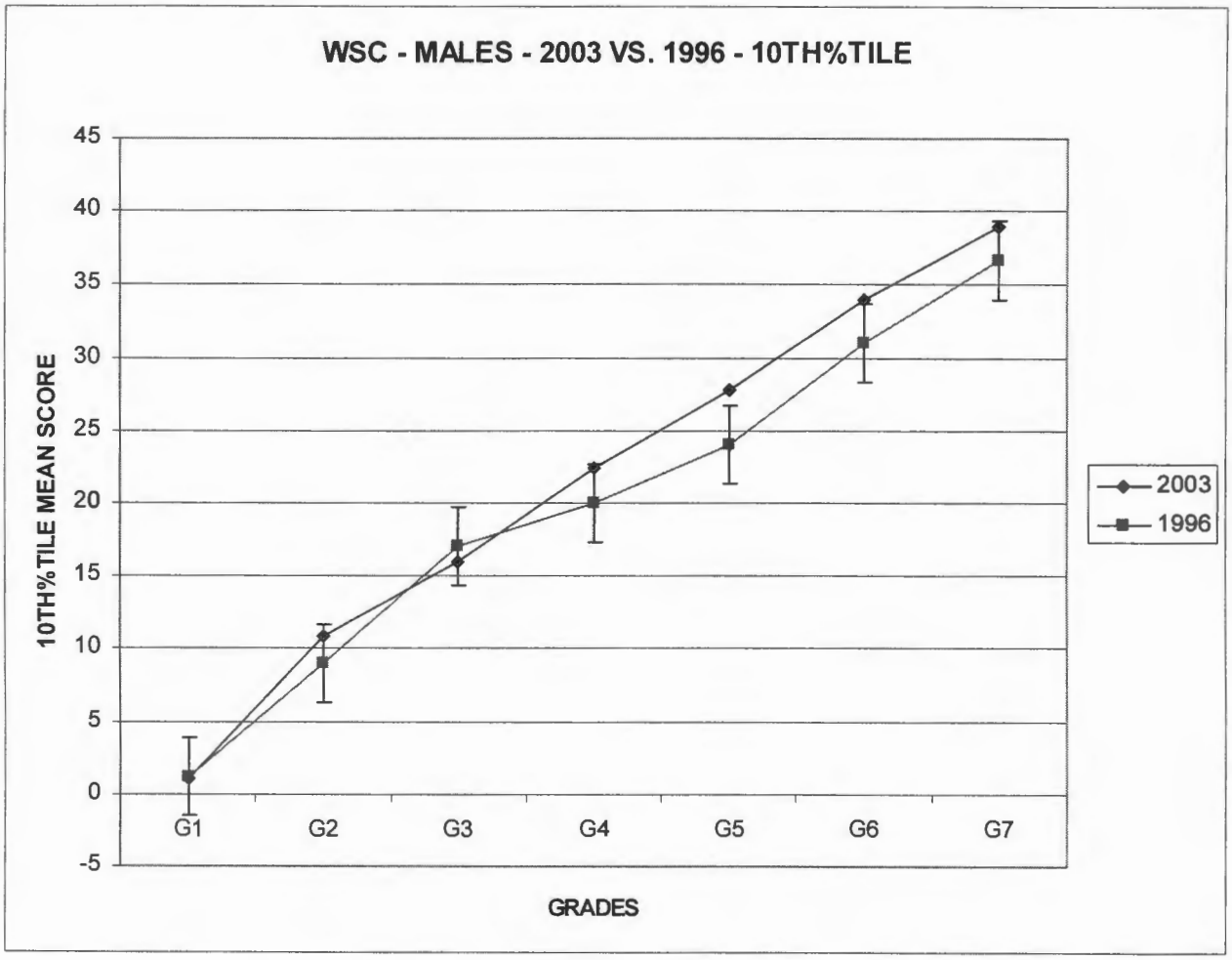

Figure 22. Grade Score Means for WSC – 10th Percentile – Males – 2003 vs. 1996

Cohen's *d* is calculated for WSC analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 46, the effect size for each grade between the two sets of data for WSC was either trivial or small.

Table 46

Effect Sizes for WSC – 10th Percentile – Males – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	P_{10}	N	P_{10}	SD	Effect Size	
One	131	1.00	111	1.20	5.26	0.04	trivial
Two	147	10.80	104	9.00	8.73	0.21	small
Three	155	16.00	126	17.00	10.44	- 0.10	trivial
Four	153	22.40	156	20.00	13.64	0.18	trivial
Five	137	27.80	160	24.00	14.84	0.26	small *
Six	162	34.00	144	31.00	14.29	0.21	small *
Seven	164	39.00	156	36.70	17.02	0.14	trivial

Total Words Written – 10th Percentile – Females – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 47 for females in Total Words Written (TWW) at the 10th percentile in 2003 and 1996. The 1996 median S.E. 1.11 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 3.56 used in Figure 23.

Table 47

Descriptive Statistics for TWW – 10th Percentile – Females – 2003 vs. 1996

Grade	(2003) P_{10}	SD	N	SE	(1996) P_{10}	SD	N	SE
One	5.00	8.45	116	0.78	4.00	7.94	165	0.62
Two	15.00	12.57	118	1.16	15.80	10.89	157	0.87
Three	24.50	13.16	124	1.18	23.00	13.03	141	1.10
Four	30.70	14.90	156	1.19	28.60	15.16	162	1.19
Five	36.00	15.60	140	1.32	40.20	16.60	151	1.35
Six	41.00	17.65	149	1.45	47.00	14.57	173	1.11
Seven	50.20	16.69	170	1.28	47.00	16.51	164	1.29

As illustrated in Figure 23, there is significant difference ($p < .05$) in the scores of TWW at the 10^{th} percentile for Females 2003 versus 1996 for Grades 5 and 6. In 2003 Grades 5 and 6 show significant regressive differences between the two data sets, while Grade 7 scores improved just inside the confidence intervals.

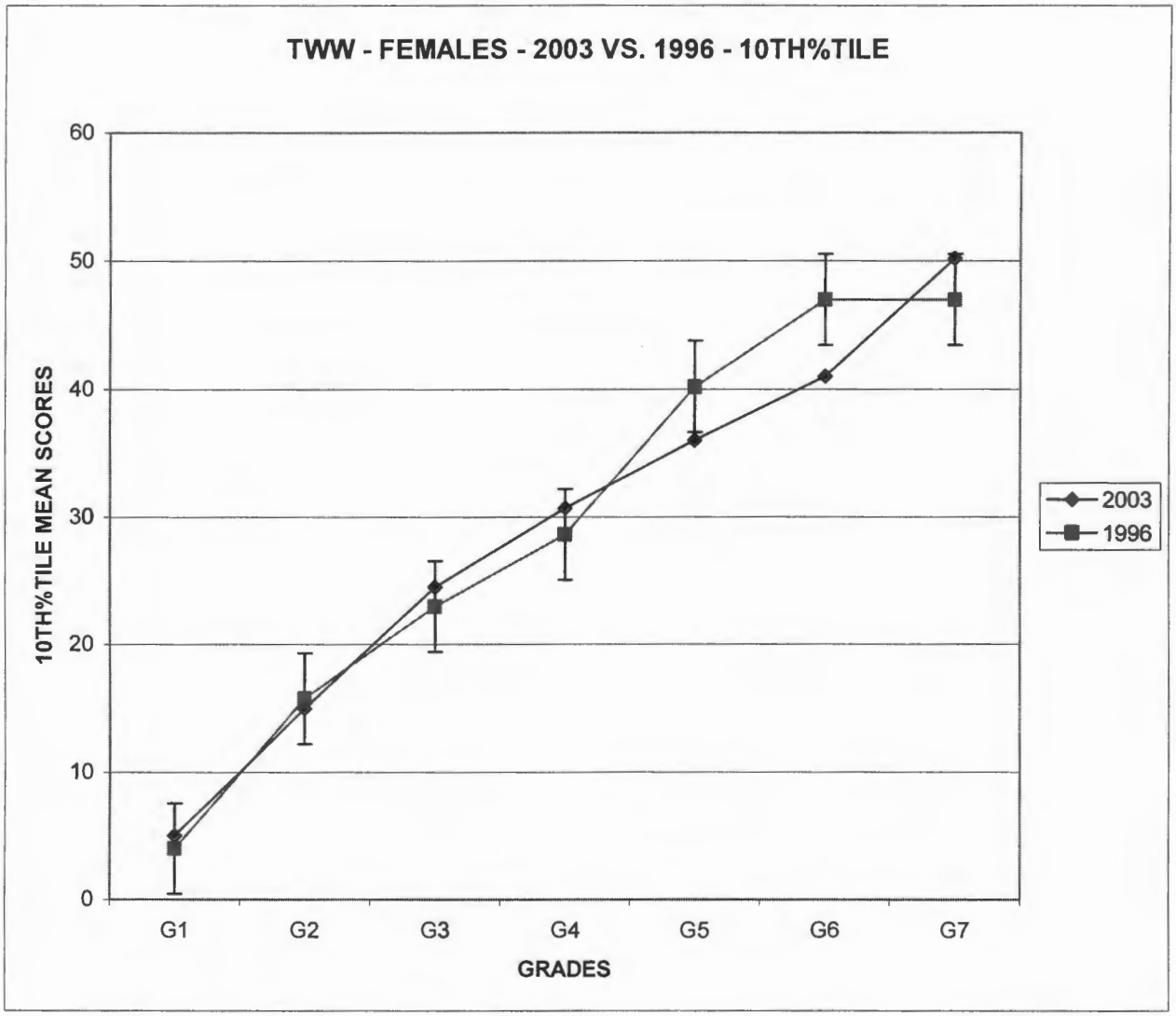

Figure 23. Grade Score Means for TWW – 10th Percentile – Females – 2003 vs. 1996

Cohen's *d* is calculated for TWW analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 48, the effect size for each grade between the two sets of data for TWW at the 10th percentile ranged from trivial to small.

Table 48

Effect Sizes for TWW – 10th Percentile – Females – 2003 vs. 1996

	(2003)		(1996)		(1996)		
Grade	N	P_{10}	N	P_{10}	SD	Effect Size	
One	116	5.00	165	4.00	7.94	0.13	trivial
Two	118	15.00	157	15.80	10.89	- 0.07	trivial
Three	124	24.50	141	23.00	13.03	0.12	trivial
Four	156	30.70	162	28.60	15.16	0.14	trivial
Five	140	36.00	151	40.20	16.60	- 0.25	small *
Six	149	41.00	173	47.00	14.57	- 0.41	small *
Seven	170	50.20	164	47.00	16.51	0.19	trivial

Total Words Written – 10th Percentile – Males – 2003 vs. 1996

The tenth percentile (P_{10}), standard deviation (SD), sample size (N), and standard error (SE) are placed in Table 49 for males in Total Words Written (TWW) at the 10th percentile in 2003 and 1996. The 1996 median S.E. 1.11 multiplied by 1.96 multiplied by 1.25, provided the 95% confidence interval of ± 2.72 used in Figure 24.

Table 49

Descriptive Statistics for TWW – 10th Percentile – Males – 2003 vs. 1996

Grade	(2003) P_{10}	SD	N	SE	(1996) P_{10}	SD	N	SE
One	3.00	7.79	131	0.68	4.00	6.81	153	0.55
Two	14.00	9.13	147	0.75	11.00	8.88	155	0.71
Three	19.00	10.90	155	0.87	20.90	10.30	178	0.77
Four	24.00	14.40	153	1.16	22.70	13.80	156	1.11
Five	30.00	15.50	137	1.32	27.40	14.30	163	1.12
Six	36.30	15.70	162	1.23	33.50	14.60	144	1.21
Seven	41.00	15.10	164	1.18	38.00	17.30	156	1.39

As illustrated in Figure 24, there is significant difference ($p < .05$) in the scores of TWW at the 10th percentile for Males 2003 versus 1996 Grades 2, 6, and 7. As can be seen in Figure 24, the confidence intervals do not overlap the scores for Grades 2, 6, and 7, signifying a definite improvement in scores between the two years for those grades. Grade 5 improvement was just inside the confidence intervals.

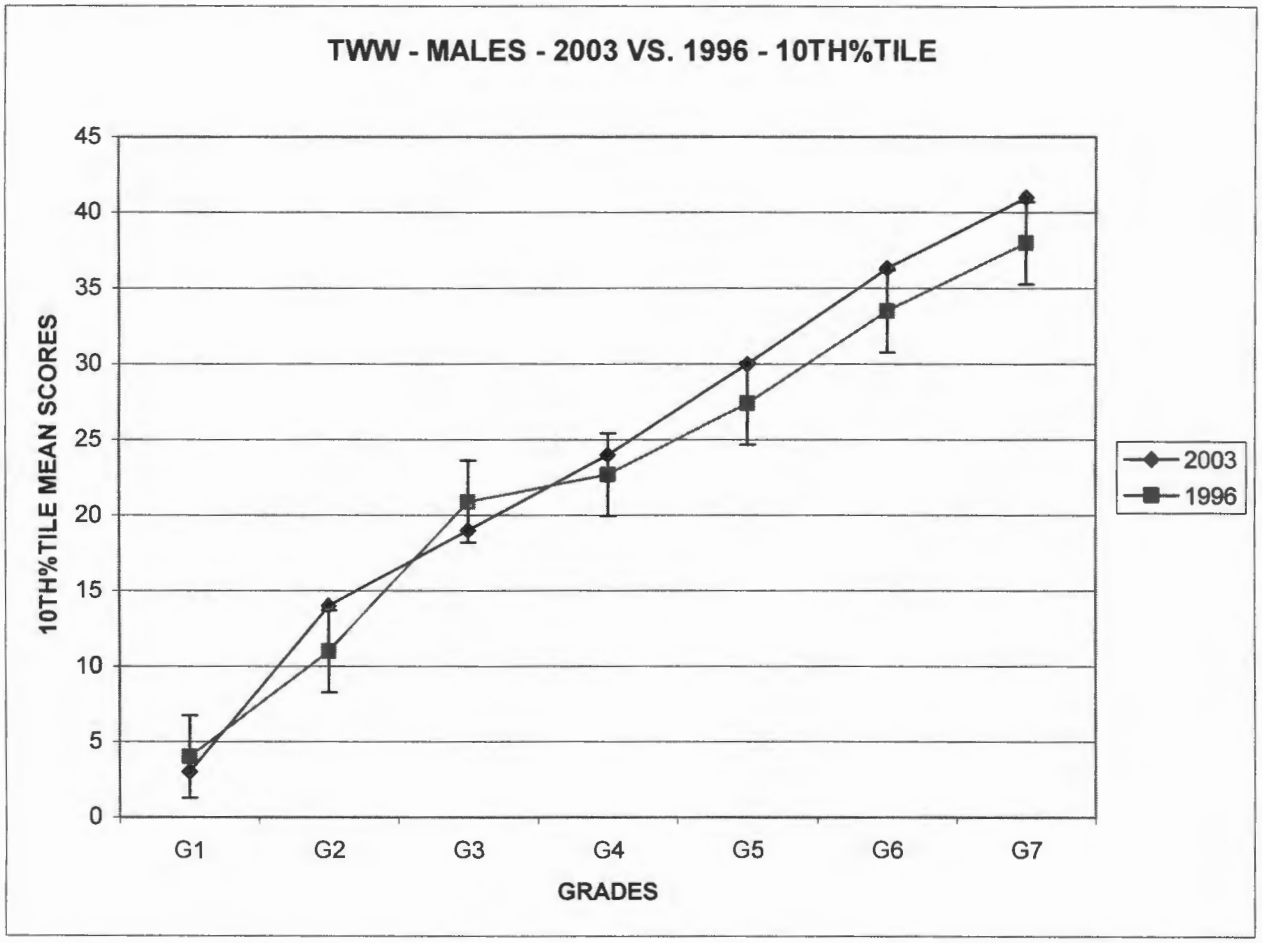

Figure 24. Grade Score Means for TWW – 10th Percentile – Males – 2003 vs. 1996

Cohen's *d* is calculated for TWW analysis at the 10th percentile for each grade for the 1996 – 2003 change. For all Cohen's *d* calculations the 1996 (baseline) standard deviation was used. Effect sizes are marked with an asterisk where there was a significant difference of $p < .05$. For all three variables the effect sizes were classified as trivial ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d \leq 0.8$), or large ($d > 0.8$). As shown in Table 50, the effect size for each grade between the two sets of data for TWW was either trivial or small.

Table 50

Effect Sizes for TWW – 10th Percentile – Males – 2003 vs. 1996

	(2003)		(1996)		(1996)	
Grade	N	P_{10}	N	P_{10}	SD	Effect Size
One	131	3.00	153	4.00	6.81	- 0.15 trivial
Two	147	14.00	155	11.00	8.88	0.34 small *
Three	155	19.00	178	20.90	10.25	- 0.19 trivial
Four	153	24.00	156	22.70	13.81	0.09 trivial
Five	137	30.00	163	27.40	14.25	0.18 trivial
Six	162	36.30	144	33.50	14.57	0.19 trivial *
Seven	164	41.00	156	38.00	17.31	0.17 trivial *

Summary of 10th Percentile – 2003 versus 1996 (Combined Genders/Female vs. Female/Male vs. Male)

In 2003 the results for combined genders at the 10th percentile in WRC, WSC, and TWW tended to have significant increases over 1996 scores in the intermediate grades (4 to 7): WRC had significant increases in Grades 3 to 7; WSC had significant increases in Grades 4, 6, and 7; and TWW had significant increases in Grades 4 and 7.

When separating the genders, it was obvious that the most gains in 2003 at the 10th percentile were with males. In WRC females showed a significant increase in Grades 3 to 6 with some regression in Grade 2, while males had significant increase in Grades 5 to 7, with marginal regression in Grade 3. In WSC females showed a significant increase in Grade 7, some regression in grade 2, and significant regression in Grades 5 and 6. Males in WSC had significant increases in Grades 6 and 7, with Grade 3 regressing.

Females in TWW significantly regressed in Grades 5 and 6, while males had significant increases in Grades 2, 6, and 7, with some regression in Grades 1 and 3.

CHAPTER FOUR: DISCUSSION AND CONCLUSION

The first series of 2003 versus 1996 comparisons of all students (combined genders) revealed that significant growth took place in nearly all grades for Words Read Correctly and primarily in the intermediate grades for Words Spelled Correctly and Total Words Written. Series two revealed that although the gap between females and males for the three variables remained statistically significant all grades effect size, males had narrowed the gap in the majority of grades. Series three revealed that although females had some increases in their 2003 scores, males had more significant increases. Males had clearly made more progress in their mean scores over the seven year span. Series four revealed that both genders had significant growth primarily in the intermediate grades at the 10^{th} percentile, with males having more growth in more grades.

All four series of data indicate that, although there was some instances regression between 2003 and 1996, the majority of mean scores had increased, with a large portion increasing significantly. The significant growth at the 10^{th} percentile for both genders revealed the importance of ongoing CBM renorming to help prevent students at this level from being over looked for support. By using outdated norms, students in need may be bypassed for the support they might have qualified for in years past.

The evidence supports the position that ongoing renorming of CBM norms will lead to continued accurate interpretation of student's scores in the future. If CBM testing is going to continue to be an indicator for support and growth, renorming should continue on a regular basis. Also, despite the fact that the gender gap has been reduced, the

challenge continues to implement strategies and interventions that will enable male students to equal female students in reading and writing fluency. Discussion within each schools growth plans may contribute to such a process.

Limitations of the study

This study did not examine school curriculum or practices that may have contributed to the changes in CBM norms. It did not interview or survey teachers to determine if they perceive differences or whether they have made conscious efforts to improve reading and writing skills for males or students at the lower ranges of achievement.

REFERENCES

Allred, R.A. (1990). Gender differences in spelling achievement in grades 1 through 6. *Journal of Educational Research, 83,* 187-193.

Cook, J. (2005) *Norming of CBM reading and writing and DIBELS instruments for school district 57.* Unpublished master's project, University of Northern British Columbia, Prince George, BC.

Deno, S.L. (1985) Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219 – 232.

Deno, S. (1992) The nature and development of curriculum-based measurement [Electronic version]. *Preventing School Failure, 36,* 5 – 10.

Dunn, R. (1991) *A comparison of curriculum – based and norm – referenced measures in the identification of reading difficulty.* Unpublished doctorial dissertation, University of British Columbia, Vancouver, BC.

Fewster, S., & MacMillan, P.D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing [Electronic version]. *Remedial and Special Education, 23,* 149-156.

Gambell, T., & Hunter, D. Surveying gender differences in Canadian school literacy [Electronic version]. *Journal of Curriculum Studies, 32,* 689-719.

Glass, G., & Hopkins, K (1996) *Statistical methods in education and psychology.* Needham Heights, MA: Allyn & Bacon.

Hedekar, L. (1997) *The effects of month of birth and gender on elementary reading and writing fluency scores using curriculum based measurement.* Unpublished master's thesis, University of Northern British Columbia, Prince George, BC.

Hedekar, L., & MacMillan, P. (in progress) *The effects of relative-age and gender on elementary reading and writing fluency scores using curriculum based measurement.* University of Northern British Columbia, Prince George, BC.

Shinn, M. (1989) *Curriculum based measurement: Assessing special children.* New York: Gilford Press.

Shinn, M., Nolet, V., & Knutson, N. (1990) Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.) *Best practices in school psychology – II.* (pp. 287 – 307). Washington, DC: National Association of School Psychologists.

Wiltshire, S. (2004) *Gender and aboriginal differences in elementary school students' CBM reading, writing, and DIBELS scores*. Unpublished master's thesis, University of Northern British Columbia, Prince George, BC.

Ysseldyke, J., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983) Generalizations from five years of research on assessment and decision making: The University of Minnesota institute. *Exceptional Education Quarterly, 4*, (1), 75 – 93.